ANNEX 81

EBC
Energy in Building and
Communities Programme

International Energy Agency

Subtask C Report

IEA Annex 81 ‘Data-Driven Smart Buildings’

Energy in Buildings and Communities
Technology Collaboration Programme

May 2025

Technology Collaboration Programme
by iea

ANNEX 81

EBC

Energy in Building and Communities Programme

International Energy Agency

IEA Annex 81 Subtask C Report

Energy in Buildings and Communities Technology Collaboration Programme

May 2025

Authors

Jin Wen, Wei Luo, Zheng O'Neill, Hicham Johra, Abbes Amira, Kamilla Heimar Andersen, Esther Borkowski, Alfonso Capozzoli, Gaurav Chaudhary, Zhelun Chen, Roberto Chiosa, Davide Coraci, Cheng Fan, Jessica Granderson, Yassine Himeur, Tianzhen Hong, Rick Kramer, Han Li, Hangxin Li; Shohei Miyata, Zoltan Nagy, Kingsley Nweye, Mohamed Ouf, Flavia de Andrade Pereira, Giuseppe Pinto, Marco Savino Piscitelli, Alberto Silvestri, Wim Zeiler

© Copyright CSIRO 2025

All property rights, including copyright, are vested in CSIRO, Operating Agent for EBC Annex 81, on behalf of the Contracting Parties of the International Energy Agency (IEA) Implementing Agreement for a Programme of Research and Development on Energy in Buildings and Communities (EBC). In particular, no part of this publication may be reproduced, stored in a retrieval system or transmitted in any form or by any means, electronic, mechanical, photocopying, recording or otherwise, without the prior written permission of CSIRO.

Published by CSIRO, PO Box 330, Newcastle, NSW 2300, Australia

Disclaimer Notice: This publication has been compiled with reasonable skill and care. However, neither CSIRO, nor the Contracting Parties of the International Energy Agency's Implementing Agreement for a Programme of Research and Development on Energy in Buildings and Communities, nor their agents, make any representation as to the adequacy or accuracy of the information contained herein, or as to its suitability for any particular application, and accept no responsibility or liability arising out of the use of this publication. The information contained herein does not supersede the requirements given in any national codes, regulations or standards, and should not be regarded as a substitute for the need to obtain specific professional advice for any particular application. EBC is a Technology Collaboration Programme (TCP) of the IEA. Views, findings and publications of the EBC TCP do not necessarily represent the views or policies of the IEA Secretariat or of all its individual member countries.

DOI:

Participating countries in the EBC TCP: Australia, Austria, Belgium, Brazil, Canada, P.R. China, Czech Republic, Denmark, Finland, France, Germany, Ireland, Italy, Japan, Republic of Korea, the Netherlands, New Zealand, Norway, Portugal, Singapore, Spain, Sweden, Switzerland, Turkey, United Kingdom and the United States of America.

Additional copies of this report may be obtained from: EBC Executive Committee Support Services Unit (ESSU), C/o AECOM Ltd, The Colmore Building, Colmore Circus Queensway, Birmingham B4 6AT, United Kingdom

www.iea-ebc.org

essu@iea-ebc.org

Preface

The International Energy Agency

The International Energy Agency (IEA) was established in 1974 within the framework of the Organisation for Economic Co-operation and Development (OECD) to implement an international energy programme. A basic aim of the IEA is to foster international co-operation among the 30 IEA participating countries and to increase energy security through energy research, development and demonstration in the fields of technologies for energy efficiency and renewable energy sources.

The IEA Energy in Buildings and Communities Programme

The IEA co-ordinates international energy research and development (R&D) activities through a comprehensive portfolio of Technology Collaboration Programmes (TCPs). The mission of the IEA Energy in Buildings and Communities (IEA EBC) TCP is to support the acceleration of the transformation of the built environment towards more energy efficient and sustainable buildings and communities, by the development and dissemination of knowledge, technologies and processes and other solutions through international collaborative research and open innovation. (Until 2013, the IEA EBC Programme was known as the IEA Energy Conservation in Buildings and Community Systems Programme, ECBCS.)

The high priority research themes in the EBC Strategic Plan 2019-2024 are based on research drivers, national programmes within the EBC participating countries, the Future Buildings Forum (FBF) Think Tank Workshop held in Singapore in October 2017 and a Strategy Planning Workshop held at the EBC Executive Committee Meeting in November 2017. The research themes represent a collective input of the Executive Committee members and Operating Agents to exploit technological and other opportunities to save energy in the buildings sector, and to remove technical obstacles to market penetration of new energy technologies, systems and processes. Future EBC collaborative research and innovation work should have its focus on these themes.

At the Strategy Planning Workshop in 2017, some 40 research themes were developed. From those 40 themes, 10 themes of special high priority have been extracted, taking into consideration a score that was given to each theme at the workshop. The 10 high priority themes can be separated in two types namely 'Objectives' and 'Means'. These two groups are distinguished for a better understanding of the different themes.

Objectives - The strategic objectives of the EBC TCP are as follows:

- reinforcing the technical and economic basis for refurbishment of existing buildings, including financing, engagement of stakeholders and promotion of co-benefits;
- improvement of planning, construction and management processes to reduce the performance gap between design stage assessments and real-world operation;
- the creation of 'low tech', robust and affordable technologies;
- the further development of energy efficient cooling in hot and humid, or dry climates, avoiding mechanical cooling if possible;
- the creation of holistic solution sets for district level systems taking into account energy grids, overall performance, business models, engagement of stakeholders, and transport energy system implications.

Means - The strategic objectives of the EBC TCP will be achieved by the means listed below:

- the creation of tools for supporting design and construction through to operations and maintenance, including building energy standards and life cycle analysis (LCA);
- benefitting from 'living labs' to provide experience of and overcome barriers to adoption of energy efficiency measures;
- improving smart control of building services technical installations, including occupant and operator interfaces;
- addressing data issues in buildings, including non-intrusive and secure data collection;
- the development of building information modelling (BIM) as a game changer, from design and construction through to operations and maintenance.

The themes in both groups can be the subject for new Annexes, but what distinguishes them is that the 'objectives' themes are final goals or solutions (or part of) for an energy efficient built environment, while the 'means' themes are instruments or enablers to reach such a goal. These themes are explained in more detail in the EBC Strategic Plan 2019-2024.

The Executive Committee

Overall control of the IEA EBC Programme is maintained by an Executive Committee, which not only monitors existing projects, but also identifies new strategic areas in which collaborative efforts may be beneficial. As the Programme is based on a contract with the IEA, the projects are legally established as Annexes to the IEA EBC Implementing Agreement. At the present time, the following projects have been initiated by the IEA EBC Executive Committee, with completed projects identified by (*) and joint projects with the IEA Solar Heating and Cooling Technology Collaboration Programme by (☼):

Annex 1: Load Energy Determination of Buildings (*)
Annex 2: Ekistics and Advanced Community Energy Systems (*)
Annex 3: Energy Conservation in Residential Buildings (*)
Annex 4: Glasgow Commercial Building Monitoring (*)
Annex 5: Air Infiltration and Ventilation Centre
Annex 6: Energy Systems and Design of Communities (*)
Annex 7: Local Government Energy Planning (*)
Annex 8: Inhabitants Behaviour with Regard to Ventilation (*)
Annex 9: Minimum Ventilation Rates (*)
Annex 10: Building HVAC System Simulation (*)
Annex 11: Energy Auditing (*)
Annex 12: Windows and Fenestration (*)
Annex 13: Energy Management in Hospitals (*)
Annex 14: Condensation and Energy (*)
Annex 15: Energy Efficiency in Schools (*)
Annex 16: BEMS 1- User Interfaces and System Integration (*)
Annex 17: BEMS 2- Evaluation and Emulation Techniques (*)
Annex 18: Demand Controlled Ventilation Systems (*)
Annex 19: Low Slope Roof Systems (*)
Annex 20: Air Flow Patterns within Buildings (*)
Annex 21: Thermal Modelling (*)
Annex 22: Energy Efficient Communities (*)
Annex 23: Multi Zone Air Flow Modelling (COMIS) (*)
Annex 24: Heat, Air and Moisture Transfer in Envelopes (*)
Annex 25: Real time HVAC Simulation (*)
Annex 26: Energy Efficient Ventilation of Large Enclosures (*)
Annex 27: Evaluation and Demonstration of Domestic Ventilation Systems (*)
Annex 28: Low Energy Cooling Systems (*)
Annex 29: ☼ Daylight in Buildings (*)
Annex 30: Bringing Simulation to Application (*)
Annex 31: Energy-Related Environmental Impact of Buildings (*)
Annex 32: Integral Building Envelope Performance Assessment (*)
Annex 33: Advanced Local Energy Planning (*)
Annex 34: Computer-Aided Evaluation of HVAC System Performance (*)
Annex 35: Design of Energy Efficient Hybrid Ventilation (HYBVENT) (*)
Annex 36: Retrofitting of Educational Buildings (*)
Annex 37: Low Exergy Systems for Heating and Cooling of Buildings (LowEx) (*)
Annex 38: ☼ Solar Sustainable Housing (*)
Annex 39: High Performance Insulation Systems (*)
Annex 40: Building Commissioning to Improve Energy Performance (*)
Annex 41: Whole Building Heat, Air and Moisture Response (MOIST-ENG) (*)
Annex 42: The Simulation of Building-Integrated Fuel Cell and Other Cogeneration Systems (FC+COGEN-SIM) (*)
Annex 43: ☼ Testing and Validation of Building Energy Simulation Tools (*)
Annex 44: Integrating Environmentally Responsive Elements in Buildings (*)
Annex 45: Energy Efficient Electric Lighting for Buildings (*)
Annex 46: Holistic Assessment Tool-kit on Energy Efficient Retrofit Measures for Government Buildings (EnERGo) (*)
Annex 47: Cost-Effective Commissioning for Existing and Low Energy Buildings (*)
Annex 48: Heat Pumping and Reversible Air Conditioning (*)
Annex 49: Low Exergy Systems for High Performance Buildings and Communities (*)
Annex 50: Prefabricated Systems for Low Energy Renovation of Residential Buildings (*)
Annex 51: Energy Efficient Communities (*)
Annex 52: ☼ Towards Net Zero Energy Solar Buildings (*)
Annex 53: Total Energy Use in Buildings: Analysis and Evaluation Methods (*)
Annex 54: Integration of Micro-Generation and Related Energy Technologies in Buildings (*)
Annex 55: Reliability of Energy Efficient Building Retrofitting - Probability Assessment of Performance and Cost (RAP-RETRO) (*)
Annex 56: Cost Effective Energy and CO2 Emissions Optimization in Building Renovation (*)
Annex 57: Evaluation of Embodied Energy and CO2 Equivalent Emissions for Building Construction (*)
Annex 58: Reliable Building Energy Performance Characterisation Based on Full Scale Dynamic Measurements (*)
Annex 59: High Temperature Cooling and Low Temperature Heating in Buildings (*)

Annex 60: New Generation Computational Tools for Building and Community Energy Systems (*)
Annex 61: Business and Technical Concepts for Deep Energy Retrofit of Public Buildings (*)
Annex 62: Ventilative Cooling (*)
Annex 63: Implementation of Energy Strategies in Communities (*)
Annex 64: LowEx Communities - Optimised Performance of Energy Supply Systems with Exergy Principles (*)
Annex 65: Long-Term Performance of Super-Insulating Materials in Building Components and Systems (*)
Annex 66: Definition and Simulation of Occupant Behavior in Buildings (*)
Annex 67: Energy Flexible Buildings (*)
Annex 68: Indoor Air Quality Design and Control in Low Energy Residential Buildings (*)
Annex 69: Strategy and Practice of Adaptive Thermal Comfort in Low Energy Buildings (*)
Annex 70: Energy Epidemiology: Analysis of Real Building Energy Use at Scale (*)
Annex 71: Building Energy Performance Assessment Based on In-situ Measurements (*)
Annex 72: Assessing Life Cycle Related Environmental Impacts Caused by Buildings (*)
Annex 73: Towards Net Zero Energy Resilient Public Communities (*)
Annex 74: Competition and Living Lab Platform (*)
Annex 75: Cost-effective Building Renovation at District Level Combining Energy Efficiency and Renewables (*)
Annex 76: ☼ Deep Renovation of Historic Buildings Towards Lowest Possible Energy Demand and CO2 Emissions (*)
Annex 77: ☼ Integrated Solutions for Daylight and Electric Lightin (*)
Annex 78: Supplementing Ventilation with Gas-phase Air Cleaning, Implementation and Energy Implications (*)
Annex 79: Occupant-Centric Building Design and Operation
Annex 80: Resilient Cooling (*)
Annex 81: Data-Driven Smart Buildings
Annex 82: Energy Flexible Buildings Towards Resilient Low Carbon Energy Systems
Annex 83: Positive Energy Districts
Annex 84: Demand Management of Buildings in Thermal Networks
Annex 85: Indirect Evaporative Cooling
Annex 86: Energy Efficient Indoor Air Quality Management in Residential Buildings
Annex 87: Energy and Indoor Environmental Quality Performance of Personalised Environmental Control Systems
Annex 88: Evaluation and Demonstration of Actual Energy Efficiency of Heat Pump Systems in Buildings
Annex 89: Ways to Implement Net-zero Whole Life Carbon Buildings
Annex 90: EBC Annex 90 / SHC Task 70 Low Carbon, High Comfort Integrated Lighting
Annex 91: Open BIM for Energy Efficient Buildings
Annex 92: Smart Materials for Energy-Efficient Heating, Cooling and IAQ Control in Residential Buildings
Annex 93: Energy Resilience of the Buildings in Remote Cold Regions
Annex 94: Validation and Verification of In-situ Building Energy Performance Measurement Techniques
Annex 95: Human-centric Building Design and Operation for a Changing Climate
Annex 96: Grid Integrated Control of Buildings

Working Group - Energy Efficiency in Educational Buildings (*)
Working Group - Indicators of Energy Efficiency in Cold Climate Buildings (*)
Working Group - Annex 36 Extension: The Energy Concept Adviser (*)
Working Group - HVAC Energy Calculation Methodologies for Non-residential Buildings (*)
Working Group - Cities and Communities
Working Group - Building Energy Codes

Summary

Subtask C of IEA EBC Annex 81 aims to provide a comprehensive understanding of the landscape of data-driven smart building applications, with a particular focus on fault detection and diagnosis (FDD) and building to grid (B2G) applications. It is believed that the current development of sensing, communication, and other computational and data advancements have opened up great opportunities to understand and operate buildings in a holistic manner to maximize a building's performance. Subtask C examines evidence from the open literature and it surveys smart-buildings experts to evaluate available Key Performance Indicators (KPIs), data-driven strategies, critical tools, frameworks, datasets, and primary market adoption barriers for FDD and B2G applications. Section 1 of this report provides a high-level overview of Subtask C as a whole, as well as a brief introduction to each of its individual activities.

Section 2 discusses the use of benchmarking algorithms. Benchmarking a building's performance, via KPIs, is critical to understanding the performance of a data-driven software application. However, it is challenging due to the complex nature of the multi-criteria assessment involved. Ideally, buildings need to be both energy-efficient and provide comfort to occupants, while also being able to provide services to energy grids in the form of demand response (energy flexibility). This activity firstly reviewed and summarized over 400 KPIs reported in the literature. Despite the abundance of proposed KPIs in the literature, challenges persist due to unclear definitions, unspecified sensor/meter data requirements, and a lack of real-life contextualization, especially at the whole-building level.

The activity then assessed the feasibility of implementing the collected KPIs using data from the existing infrastructure of five office buildings located in the Netherlands and Switzerland. A survey was also performed to seek stakeholders' opinions on the significance of various KPIs. The case studies and stakeholders' survey studies clearly revealed that there is a misalignment among KPI definitions in the literature, the data that is readily available from existing building management systems (BMS), and stakeholders' needs. Further research is required to contextualize KPIs across diverse application scenarios while taking stakeholder perspectives into consideration.

Section 3 focuses on applications that use building operational data and building models to identify faults and isolate their root causes in an automated way. Compared with traditional expert-knowledge/rule-based FDD methods, which are typically seen in current market-available FDD products, data-driven FDD methods require little or no *a-priori* knowledge, and hence can have higher accuracy and autonomy with lower cost. This activity began by reviewing existing literature on data-driven FDD, including its definition, framework, methods, applications, and evaluation criteria. The literature review identified many promising methods and frameworks for implementing data-driven FDD, and for their application in different building systems. A comprehensive evaluation was made of metrics developed for data-driven FDD performance evaluation.

The literature review also found that there is generally a lack of high-fidelity data suitable for FDD development and evaluation (especially that from real building systems). Therefore, to further promote market adoption of data-driven FDD methods, both (i) a publicly available data repository for FDD development and evaluation, and (ii) a list of market-available FDD software, were compiled. The results are summarized in Section 3.3.

Based on our survey studies and the literature review, a roadmap was developed to guide industry stakeholders, including building owners, end users (operators, facility managers, etc.) and building technology providers. It aims to communicate the barriers and possible solutions for growing the adoption of FDD applications. Barriers were identified under five categories, which include:

- Economic and business: Costs and benefits for end-users and/or business limitations.
- Technological: Technical knowledge, interoperability, digital infrastructure and/or data management.
- User-related: User experience, interfaces and/or misunderstandings.
- Regulatory: Policies, GDPR and/or cybersecurity.

- Social and societal: Cultural, community and stakeholders, benefits for society and/or environmental sustainability.

The barriers and their potential solutions were identified for each stakeholder across the FDD implementation lifecycle, which includes the FDD solution development stage, deployment stage, data management stage, and fault analysis and solution handling stage. For example, a technological barrier during the FDD solution development stage is the challenge to integrate the FDD solution with the BMS. A potential solution is to adopt an ontology and semantic data management principles. Another example is social barriers during the FDD implementation stage, including both public skepticism about data privacy, and possible impacts on employment. Potential solutions include public awareness campaigns on the benefits of FDD and training programs. This roadmap helps to guide industry stakeholders through the FDD ecosystem and to provide insights and direction settings for future research and development.

Section 4 contributes to standardizing B2G service assessment. A literature review was performed on data-driven KPIs, relevant to the assessment of building energy flexibility during the operational phase. 81 KPIs were identified. The review highlighted that the two main constraints in quantifying energy flexibility are 1) many of the KPIs are baseline dependent and there is a lack of robust data-driven approaches for generating these baseline load profiles (i.e. when demand response is not activated); and 2) there is a lack of KPIs that can be computed without reference to a baseline or reference scenario inputs (baseline-free KPIs). It was also found that most studies (65%) were conducted using numerical simulations.

To develop, study, test and benchmark B2G services at scale, 16 open-access building energy flexibility datasets were gathered. Section 4.3 describes these datasets and explores their use for B2G application development and testing. Currently, no standardized method exists for evaluating load flexibility from B2G services. To address this gap, an open-source toolbox in the form of a Python package is reported in Section 4.4. It offers a data-driven approach for B2G service assessment using collected KPIs. An online platform was also developed, which is discussed in Section 4.5. The online platform features an ontology explorer to guide users through a new ‘EFOnt’ ontology (meta data schema). It helps to identify the most suitable KPI for their energy flexibility assessment needs. Users can then choose to utilize the automatically collected dataset or upload their own dataset to assess their B2G services. The collected KPIs, open datasets, and developed toolkits will facilitate benchmarking and understanding of B2G services, promoting greater market adoption in the future.

Table of contents

Figures

Tables

Abbreviations

Abbreviations	Meaning
AFDD	Automated Fault Detection and Diagnosis
AHU	Air Handling Unit
ASHRAE	American Society of Heating Refrigeration and Airconditioning Engineers
AUC	Area Under the Curve
B2G	Building to Grid
BMS	Building Management System
BN	Bayesian Networks
CDR	Correct Diagnosis Rate
CORAL	Correlation Alignment
CV-RMSE	Coefficient of Variation of Root Mean Square Error
DNN	Deep Neural Network
DT	Decision Tree
DTW	Dynamic Time Warping
EC	Energy Consumption per square meter
FCU	Fan Coil Unit
FDD	Fault Detection and Diagnosis
FF	Flexibility Factor
FNR	False Negative Rate
FPU	Fan Power Unit
GAN	Generative Adversarial Network
HMM	Hidden Markov Model
HVAC	Heating Ventilating and Air Conditioning
IEA	International Energy Agency
JS	Jensen–Shannon
KPI	Key Performance Indicator
LF	Load Factor

LIME	Local Interpretable Model-agnostic Explanations
LLM	Large Language Model
MAE	Mean Absolute Error
MAPE	Mean Absolute Percentage Error
MBE	Mean Bias Error
MK-MMD	Multiple Kernels MMD
ML	Machine Learning
MMD	Maximum Mean Discrepancy
MSE	Mean Squared Error
NN	Neural Networks
NRMSE	Normalized Root Mean Squared Error
PCA	Principal Component Analysis
PCO_2	Percentage of CO_2 exceeding the threshold
PMV	Predicted Mean Vote
PPD	Predictive Percentage of Dissatisfaction
RES	Renewable Energy Sources
RMSE	Root Mean Square Error
ROC	Receiver Operator Characteristic
ROI	Return On Investment
RTU	Roof Top Unit
SHAP	SHapley Additive exPlanationsis
SRI	Smart Readiness Indicator
SVM	Support Vector Machine
SVR	Support Vector Regression
TL	Transfer Learning
TPR	True Positive Rate
VAV	Variable Air Volume
VCC	Vapor Compression Cycle
WD	Wasserstein Distances

1. Task C overview

Having access to diverse high-quality data (which come from sensor measurements, digital twin models, and other data sources), opens up a range of new possibilities for understanding and optimizing building performance and operation from a holistic perspective. Holistic here refers to simultaneously taking into consideration multiple key performance indicators (energy, comfort/health, maintenance, load flexibility, etc.), and data from multiple sources.

This subtask focuses on understanding the state-of-the-art of key data-driven software applications, including their data-driven strategies, and identifying primary barriers for these applications to be adopted by the market. It also aims to address information barriers by establishing Key Performance Indicators (KPIs) and frameworks for benchmarking of buildings and software applications (particularly fault detection and grid enabled buildings).

Subtask C summarizes critical tools, frameworks, and information for software developers to develop 'applications' that can be commercialized for fault detection, reducing energy consumption in buildings, and coordinating building energy demand to achieve additional electricity system benefits.

The work of Subtask C is organized into the following three activities

- C.1 Benchmarking algorithms. Benchmarking a building's performance via KPIs is critical to understanding the performance of a data-driven software application. A large quantity of data-driven KPIs have been reported in the literature, which can be categorized into the following four aspects: occupant-centric, building smartness, building energy and maintenance, and building-grid interaction. This activity firstly reviews and summarizes literature reported KPIs, including their algorithms (Section 2.2). The activity then comprehensively assesses the feasibility of implementing the collected KPIs using data from the existing infrastructure of five office buildings located in the Netherlands and Switzerland (Section 2.3). A survey was carried out to seek stakeholders' opinions on the significance of various KPIs, as outlined in Section 2.4.
  The identified benchmarking algorithms and KPI values could be used for evaluating the effectiveness of different data-driven software applications.

- C.2 Automated fault detection, diagnostics and recommissioning. This activity focuses on applications that use building operational data and building models to identify faults and isolate their root causes in an automated way. This activity firstly reviews existing literature on data-driven fault detection and diagnostics (FDD), including its framework, methods, applications, and evaluation criteria (Section 3.2). A publicly available data repository for FDD development and evaluation, as well as a survey of market-available FDD software is summarized in Section 3.3. A roadmap is also developed (Section 3.4) to guide industry stakeholders to understand the barriers and possible solutions for growing the adoption of FDD applications.

- C.3 Building2Grid. This activity focuses on applications for orchestrating flexible demand from buildings in a way that can dampen the dynamics of the electricity grid. In collaboration with activity C1, activity C3 firstly reviews data-driven KPIs for the operational phase of buildings performing Building2Grid services and demand response (Section 4.2). In order to develop, study, test and benchmark B2G services at scale, Activity C3 then collects open-access building energy flexibility datasets (in the form of time series data). Section 4.3 describes the collected datasets and analyzes their application for B2G application development and testing. At the moment, there is no standard way to assess B2G services. To address this limitation, activity C3 has developed an open-source toolbox in the form of a Python package for data-driven assessment of B2G services (Section 4.4). An online platform is also developed for analysis of B2G services (Section 4.5).

2. Activity C1

2.1 Introduction

Benchmarking building operational performance is a pivotal step toward achieving energy-efficient, healthy, and comfortable buildings. However, it is challenging due to the complex nature of the multi-criteria assessment involved. Ideally, buildings need to be both energy-efficient and provide comfort to occupants, while also being able to provide services to energy grids in the form of demand response (energy flexibility). KPIs can represent critical pieces of actionable information and help to evaluate and track if buildings meet their objectives [1]. Therefore, the research team in activity C1 focused on investigating the benchmarking algorithms for building operational performance with specific emphasis on the KPIs. Given that a good KPI should be accessible, quantifiable, and actionable [2], the work of activity C1 is structured in a way that addresses three key research questions (Figure 1):

- Research question 1: What are the KPIs in existing buildings and energy-related literature?
- Research question 2: What KPIs can be implemented within the existing infrastructure of current buildings?
- Research question 3: Which KPIs are important to the building stakeholders?

The Smart Readiness Indicator (SRI) is an assessment scheme in the EU, that assesses the readiness of building smart services. It includes three key functionalities: 1) Respond to the occupants' needs, 2) Respond to the grid's demand, and 3) Maintain building energy efficiency and operation. The KPI framework developed in activity C1 draws inspiration from the Smart Readiness Indicator framework [3] incorporating an additional category to evaluate the efficacy of smart technology (Figure 1).

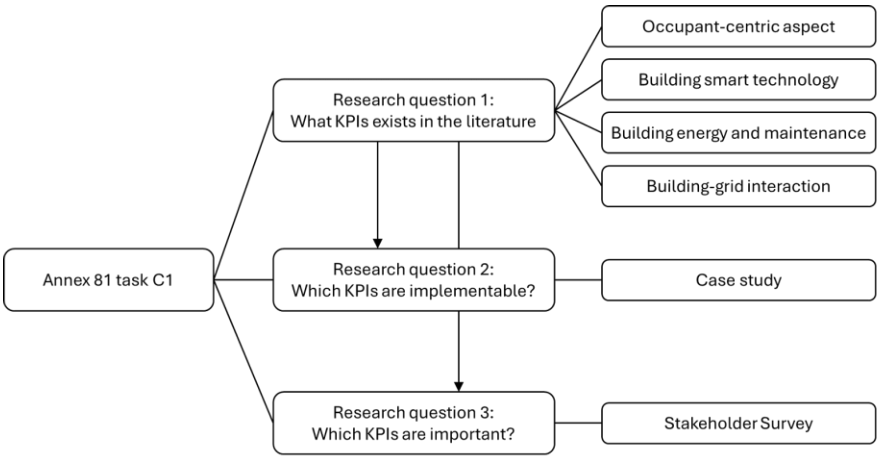

Figure 1 Research framework of Annex 81 Activity C1.

Activity C1 first provides an overview of the KPIs found in the literature based on the proposed KPI framework (research question 1). These KPIs are then tested using real data from case study buildings (research question 2). Finally, a survey is conducted to collect stakeholders' opinions on building performance priorities (research question 3).

2.2 Literature review summary

The KPIs identified from the literature can be categorized into four categories, namely, occupant-centric KPIs, interoperability KPIs, building energy saving and maintenance KPIs, and energy flexibility KPIs.

2.2.1 Occupant-centric KPIs

Occupants, as the primary beneficiaries of building services, wield considerable influence on the energy performance of buildings. Simultaneously, building operations shape occupants' comfort, productivity, and overall well-being. The increased focus on occupant health and comfort, exacerbated by the challenges posed by COVID-19, has magnified the importance of occupant-centric considerations. The review in this section categorizes occupant-centric KPIs into three main areas: those related to occupants' interaction with building systems, those based on indoor/outdoor environmental parameters, and those related to occupants' subjective feedback. Notably, the impracticality of measuring human-building interactions during the pre-occupancy phase necessitates the post-occupancy calculation of those KPIs. Although simulation-based assessments are an alternative, their reliance on simplified assumptions poses challenges for accurately quantifying real building performance from an occupant-centric perspective.

Drawing from the building life cycle, stakeholders relevant to occupant-centric KPI mostly include building designers, occupants, building managers, and building owners. From these stakeholders' perspectives, the review identified 22 thermal KPIs, 11 air quality KPIs, 10 acoustic KPIs, and 17 visual-lighting KPIs, quantifying occupants' comfort, health, productivity, and well-being [4]. The comfort-related KPIs are the most popular ones. Dry air temperature, CO_2, sound level and illuminance are the most required input data for thermal, air quality, acoustic, and visual KPIs, respectively. Nevertheless, although KPI calculation formulas exist, challenges persist due to unclear definitions and a lack of specified sensor/meter data requirements, particularly on the whole building level. Further work is needed to specify the methodology to compute these occupant-centric KPIs at a larger scale.

2.2.2 Interoperability KPIs

"Interoperability is the ability of two or more systems or components to exchange data and use information" [5]. Interoperability enables communication between different building systems and between building systems and the energy grids - a crucial aspect of smart buildings technology. The significance of interoperability testing in achieving seamless integration is widely acknowledged [6]. Despite its recognized importance, interoperability testing lacks common specifications, and the absence of universally accepted quantifiable KPIs in building domain testing is notable. There are a few methodological approaches for interoperability assessment in the smart grid domain [6]. For example, Ford et al. [7] developed an i-Score methodology to assess the interoperability of networks of systems. This methodology abstracts the systems as an architecture framework that describes how these systems work. Based on the architecture data, it employs graph optimization, and interoperability theory to offer a comprehensive assessment of interoperability. Van Amelsvoort et al. [8] then adapts this i-Score methodology for interoperability testing in the smart grid domain. However, the prevailing practice of devising adhoc interoperability testing procedures, without embracing well-structured methodological approaches, can result in issues such as irreproducibility, subpar quality, prolonged development times, and increased costs [6]. Despite the evident importance of interoperability, progress in initiatives to enhance the current situation is sluggish, making it premature to operationalize within the framework of Annex 81.

2.2.3 Transfer learning KPIs

Transfer Learning (TL) is a powerful technique in Machine Learning (ML) where a model trained on a specific task (i.e., source task, or a source building) within a particular domain can be applied to a new task (i.e., target task, or a target building) that shares similarities with the original task, whether within the same domain or across different domains. In the context of smart buildings technology, implementing a transfer learning strategy can improve model performance, reduce the model computation time, and lower the cost of deploying smart algorithms. This could be for use cases such as load prediction, occupancy detection and activity recognition, building dynamics, advanced control systems and fault detection and diagnosis. The traditional TL process includes 1) identifying the best source domain (building) using similarity metrics; 2) applying TL solutions; 3) assessing the TL performance.

The different nature of data in the built environment has led to several methods to quantify building similarity based on specific applications. From the analysis of different applications in the built environment, two different approaches have been identified: semantic approach and data-based approach. The semantic approach uses features, metadata, and semantics to study the similarities between two buildings, while the data-based approach analyses the datasets available, trying to assess similarities between the source and the target datasets, using both features and time series.

The transfer learning solutions can be categorized into four types [9]: 1) Instance-based TL, where knowledge is transferred by utilizing data from similar environments or tasks to improve the target domain; 2) Feature-representation TL, where knowledge is transferred through the learned representation of features; 3) Model-parameter TL, where knowledge is transferred by sharing model parameters or their distributions between the source and target domains; 4) Relational-knowledge TL, where knowledge transfer focuses on the relationships or interactions between entities in a dataset, specific to the relational structure of the data.

The assessment of TL performance requires the definition of several metrics to assess building similarity (i.e., domain similarity) and machine learning performance, that can be employed to compute KPIs that quantify TL advantages in terms of performance, speed, data requirements and reliability. A number of KPIs have been introduced in [10] to quantify the performance of TL for building applications such as jumpstart, transfer ratio, asymptotic performance, time to threshold, performance with fixed number of epochs, performance sensitivity, necessary knowledge amount, and necessary knowledge quality. However, each KPI needs to be contextualized in the framework of the TL application. A brief overview of metrics employed to quantify similarity between different buildings and of KPIs for TL performance, is provided below for each main building energy management application.

Load prediction: Adopting machine learning techniques for load prediction is still facing many challenges in real applications. In practice, limited data due to lack of monitoring infrastructure or time of data accumulation is a major barrier. Transfer learning is an effective approach to solve the data limitation problem in building load prediction and has attracted growing research interests. Several common KPIs were used to evaluate the performance of prediction models [11] and the consequent improvement when a TL framework is implemented. These KPIs include RMSE (Root Mean Square Error) improvement [12], CV-RMSE (Coefficient of Variation of Root Mean Square Error) improvement [13] and MAPE (Mean Absolute Percentage Error) improvement [14]. Moreover, various techniques have been used to assess building similarity, as Dynamic Time Warping (DTW) [15], Similarity Measurement Index [14], Maximum Mean Discrepancy (MMD) [16] and Mahalanobis Distance [17].

Occupancy detection and activity recognition: In the framework of occupancy detection, transfer learning could be useful in addressing some of the challenges faced when developing occupancy detection models. These challenges border on sensing infrastructure limitations [18], labeled data collection [19], model quality as well as generalization [20], and model explainability [21]. The evaluation metrics used to quantify the occupancy detection model performance are distinguished by the modeling problem under study. The works concerned with occupancy count prediction utilized metrics such as Mean Squared Error (MSE), Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), Normalized Root Mean Squared Error (NRMSE) and Mean Bias Error (MBE) to infer their model quality [22]. On the other hand, works that focused on occupancy presence detection (binary classification problem) used either accuracy, F1-score or both metrics to report model performance [23]. To quantify the performance of the TL improvement in activity recognition, the literature review revealed that the most commonly used metrics are Recall, Precision, and F1-Score [24]. An advantage of these metrics is that they are not affected by the distribution of the different classes and usually provide more accurate values for frequently occurring activities like sleep. Other performance metrics for binary classification for occupancy detection are Balanced Accuracy, Matthews Correlation Coefficient, Area under the Receiver Operating Characteristic Curve, and manual inspection of the confusion matrix [25].

Building dynamics: The application of transfer learning for building thermal dynamic models has proven to speed up the training process of the data-driven models in data-scarcity conditions and increase their accuracy and performances. The main machine learning metrics used to quantify the performance of building dynamics applications are RMSE, MAE and MAPE [9, 26]. These metrics are employed to quantify the effectiveness of TL on building thermal dynamic performances by employing KPIs as performance improvement ratio (representing the percentage improvement of using TL over classical machine learning), performance with a fixed number of epochs (comparing the performance of TL and no-TL applications with a fixed number of epochs, where the performance difference with 1 epoch represents the jumpstart) and time to the threshold (representing the time in terms of training epochs needed to reach a certain performance).

Control systems: The implementation of TL in building control can offer numerous advantages. For instance, it simplifies the transfer of information between advanced controllers, streamlining the application of algorithms that are often customized for specific control problems. Additionally, using TL scaling the implementation of such algorithms in buildings with limited historical data [26] could significantly reduce the training time required for data-driven control strategies to achieve near-optimal control policies. Several metrics to measure the performance of TL for advanced controllers are established according to the objective functions. In this framework, controller performances are mainly related to energy consumption or cost [27] and comfort conditions [28]. Energy consumption or energy cost refers to the energy/cost reduction allowed by the implementation of an advanced pre-trained controller transferred from a source building compared to a baseline controller (rule-based controller or advanced controller trained from scratch). Moreover, from the analysis of the literature reviews, the main KPIs used to quantify the effectiveness of TL for building control are time to threshold, performance with a fixed number of epochs [27], time to reach the performance of a controller not transferred and trained from scratch (representing the time in terms of training epochs needed to reach certain performance) [29]. The definition of robust metrics and KPIs to quantify the similarity between source and target buildings is one of the main challenges related to the application of TL for building control applications.

Fault Detection and Diagnosis (FDD): The acquisition of comprehensive datasets fully representative of the normal and faulty conditions of a system is essential to pursue a reliable data-driven approach to FDD but at the same time is laborious, time-consuming, and costly. In addition, a significant discrepancy between training and testing datasets due to load change and different operation modes can seriously affect their performance in promptly detecting and diagnosing faults during system operation [30]. Researchers worldwide identified transfer learning as a valuable approach that can be employed to overcome these challenges and to enhance the applicability of data-driven FDD processes. More discussions specifically on FDD are provided in Section 3.2. According to [31], [16], and [32], the criteria that are proven to be also successful in FDD tasks, to assess discrepancy between distribution similarity of transferable features, include maximum mean discrepancy (MMD), Kullback–Leibler divergence, multiple kernels MMD (MK-MMD), Jensen–Shannon (JS) divergence, Correlation Alignment (CORAL), Euclidean distance, Wasserstein Distances (WD), OT-embedded joint distribution similarity measure (OT-JDSM). Given that the diagnosis problem in FDD is mostly defined as a classification task, the employed KPIs to evaluate the beneficial effect of TL are related to Accuracy, Precision, Recall and F-measure improvement degree as reported in [33].

2.2.4 Building energy saving and maintenance KPIs

Energy performance indicators are normally integrated into rating and certification systems based on building energy codes and standards. Li et al. [34] summarized the most common energy performance indicators at the building level and introduced a set of system-level KPIs that include four major end-use systems and their eleven subsystems.

Building maintenance encompasses a range of activities aimed at preserving and repairing the functionality, safety, and aesthetics of a building and its components. Maintenance costs can account for up to 65% of annual facility management costs [35]. [36] introduces three types of maintenance: Improvement maintenance, predictive maintenance, and corrective maintenance. Based on these maintenance types, the performed review has comprehensively listed the maintenance KPIs incorporated in the [37]. Maintenance KPIs are classified into eight groups: physical asset management (20 KPIs), information communication technologies (20 KPIs), health safety environments (22 KPIs), maintenance management (22 KPIs), people competence (20 KPIs), maintenance engineering (19 KPIs), organization and support (30 KPIs), and administration and supply (29 KPIs). In addition, activity C2 specifically reviewed the maintenance KPIs related to Fault Detection and Diagnosis. The review [38] classifies the FDD KPIs into three categories: general evaluation metrics for FDD applications (8 KPIs), evaluation metrics for data-driven classification problems (5 KPIs), and statistical significance tests that assist the evaluation of classification problems (5 tests).

2.2.5 Energy flexibility KPIs

A recent literature review in Activity C3 (Section 4.2) identified 29 generic KPIs and 48 data-driven KPIs for assessing demand response and building energy flexibility [39]. These KPIs can be categorized into 12 distinct groups: power peak shedding, energy/average power load shedding, peak power/energy rebound, valley filling, load shifting, demand profile reshaping, energy storage capability, demand response energy efficiency, demand response costs/savings, demand response emission/environmental impact, grid interaction, and impact on indoor environment quality. These KPIs usually have low complexity, but most of them (81%) require a baseline (scenario without demand response) to be calculated. The most popular of these KPIs are related to the energy efficiency of a demand response action, the load shifting capacity (typically from high-price periods to low-price periods), and the peak power shedding. However, the popularity of these KPIs in the research community does not necessarily reflect their applicability and usefulness for the industry and other demand response stakeholders. For instance, Johra et al. [11] showed that there is a clear mismatch between the data that is typically available from current BMS datasets and the data requirements to compute these KPIs. In addition, a discussion with the different stakeholders is necessary to identify what assessment method is the most appropriate for their specific use.

2.2.6 Summary

In summary, the literature overview has collected 60 occupant-centric KPIs, 40 KPIs for transfer learning, 274 KPIs for building energy and maintenance, and 77 KPIs for building-grid interaction, resulting in a total of 451 KPIs (Figure 2).

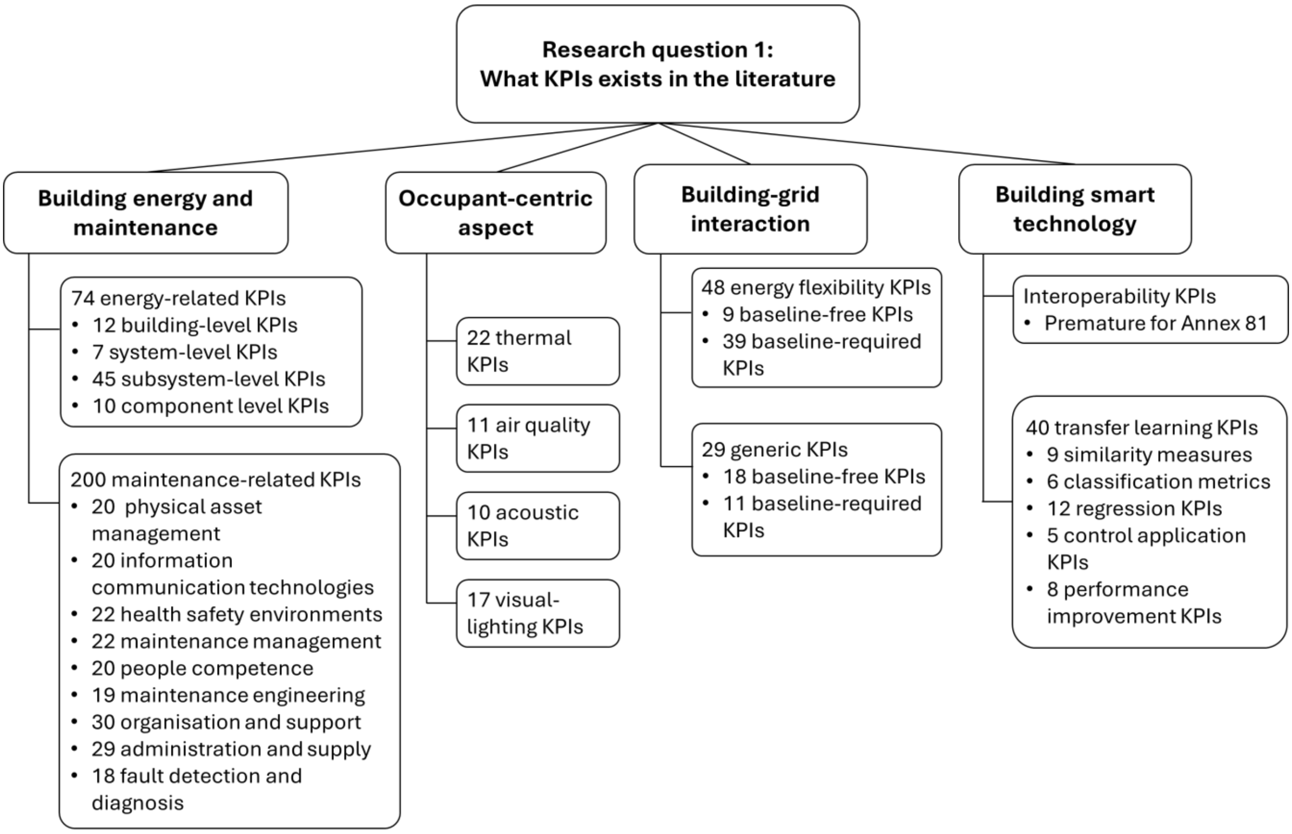

Figure 2 Overview of the collected KPIs.

2.3 Case study

This section presents a comprehensive evaluation of previously gathered KPIs across five office buildings, with four located in the Netherlands and one in Switzerland. The analysis uses historical BMS data from 2022 to ascertain the feasibility of computing various KPIs, focusing particularly on occupant-centric and energy flexibility metrics. The findings underscore challenges associated with data availability for KPI computations. Thermal comfort KPIs are found to be the most readily calculable among occupant-centric KPIs, while those related to building lighting and acoustics present significant challenges if using BMS data. Similarly, within energy flexibility KPIs, only those dependent on total energy demand are generally calculable. On average, only approximately one-fourth of the collected KPIs can be reliably calculated for the case study buildings (Figure 3).

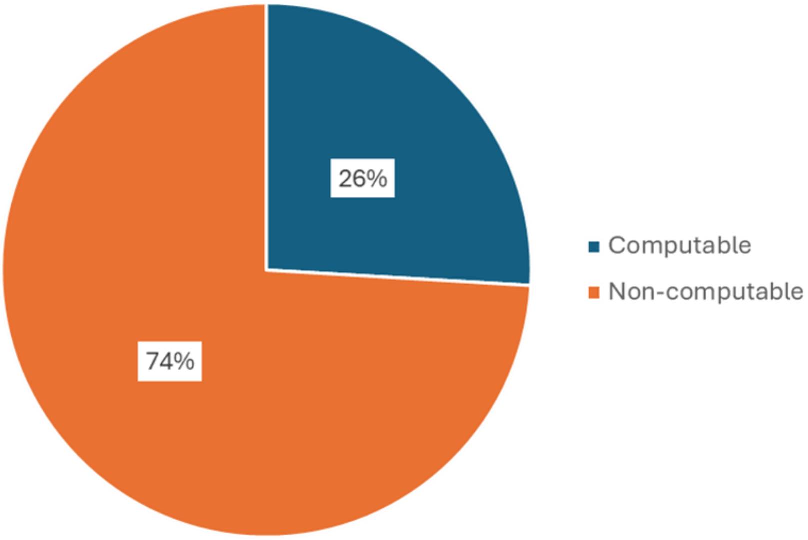

Figure 3 Percentage of the computable KPIs using historical BMS data of case buildings.

Furthermore, a detailed assessment of six KPIs, namely Predicted Mean Vote (PMV), Predictive Percentage of Dissatisfaction (PPD), Percentage of CO_2 exceeding the threshold (PCO_2), Energy Consumption per square meter (EC), Flexibility Factor (FF), and Load Factor (LF), was conducted to gauge the annual performance of the case study buildings. While these KPIs can be computed, their definitions lack consideration for the complexity of real-world building scenarios, introducing ambiguity and limiting reliability in calculations. For instance, the PMV index relies on indoor air temperature as an input parameter. In the case study buildings, one building has forty indoor air temperature sensors distributed in various rooms, while another only has one indoor air temperature sensor. This variation in the number and placement of indoor air temperature sensors can lead to bias in the PMV calculation when comparing the thermal performance of two buildings. Several key considerations are highlighted:

- Input data quality: KPI definition should specify input data quality, such as sensor accuracy, sampling frequency, maximum amount of missing data points, and outlier management.
- Spatial factors: KPIs should account for the spatial distribution of sensors to ensure a representative measurement of the entire space.
- Temporal factors: KPIs should describe the time resolution for calculations, ranging from annual to sub-hourly intervals.
- Data aggregation factor: The data aggregation factor should be addressed, indicating how data is aggregated in the temporal dimension, from lower-level (e.g., minute-level sensor data) to higher-level intervals for KPI calculations and in the spatial dimension, how to aggregate sensors over large buildings with distinct thermal zones. Different aggregation methods may affect the calculations.

Analysis of collected data indicates that spatial factors are the most influential for PMV, PPD, and PCO_2 calculations, while temporal factors and data aggregation factors play a more critical role in FF and LF computations. Importantly, the significance of these considerations depends on the specific KPIs, building characteristics, performance goals, sensor technologies, and their interplay. This again underscores the need for further research to standardize KPIs, ensuring a reliable benchmarking process for assessing building performance in practical applications.

2.4 Stakeholder survey on KPI

Numerous KPIs could be generated and calculated in the previous sections; however, not all of them are important to the stakeholders. To address this, a survey was conducted, seeking input from stakeholders on essential aspects of building operational performance. The survey was designed based on the proposed KPI framework, incorporating three building performance goals: 1) to improve buildings' energy saving and operation, e.g., energy efficiency, operational cost, environmental impact, and maintenance. 2) to satisfy occupants' needs, e.g., comfort, health, well-being, and convenience. 3) to satisfy the grid's requirements and provide building-to-grid services, e.g., grid stability and demand response. Each general goal was further subdivided into four sub-performance/technical aspects. The survey employed the Analytic Hierarchy Process (AHP) to gauge stakeholders' opinions on the relative importance of two performance aspects and to calculate their corresponding weights.

A total of 137 stakeholders received the questionnaire, with 65 stakeholders (47.4% response rate, predominantly building managers) completing the survey. The results indicate that stakeholders typically prioritize occupants' needs the most, followed by the building's energy efficiency and operation, and exhibit the least concern for the grid's requirements. Within the occupants' needs category, occupant health emerged as the most important aspect and sub-aspects like mitigating respiratory disease transmission, followed by comfort. For building operations, stakeholders considered the downtime of the building system as the most critical consideration, while operational cost ranked as the least important. In contrast, for building energy flexibility, all technical aspects held similar importance, encompassing power peak shedding, energy/average power load shedding, peak power/energy rebound, valley filling, load shifting, demand profile reshaping, and energy storage capability. However, the study also unearthed notable variations in priority among individual stakeholders. Specifically, only 52% ranked occupants' needs highest, while a smaller fraction (14%) deemed the grid's requirements their foremost concern (Figure 4). This may be caused by many factors, such as stakeholder type, the building functions, policy, and country.

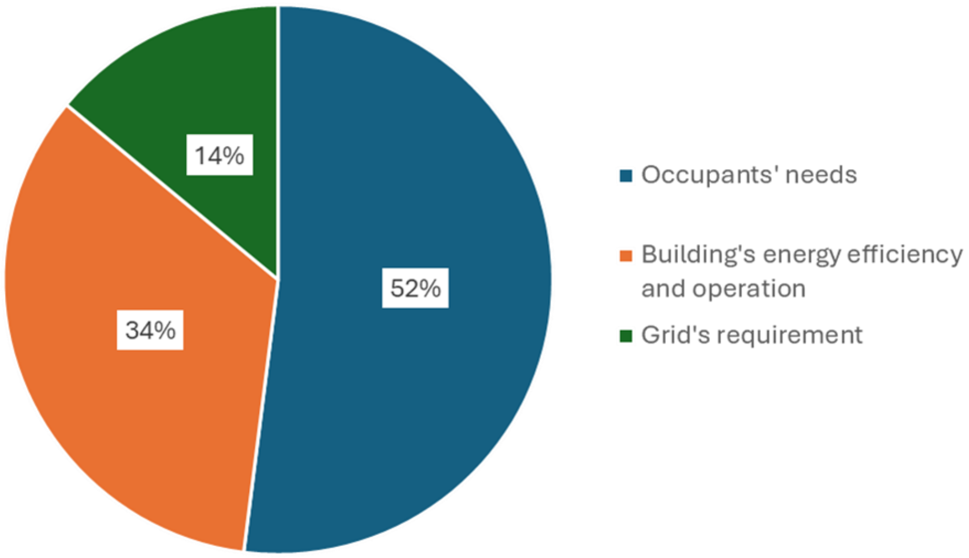

Figure 4 Percentage of stakeholders who prioritize one building performance aspect.

3. Activity C2

3.1 Introduction

With the wide adoption of building management systems, and the advancement of data, sensing, and machine learning techniques, data-driven FDD for building heating, ventilation, and air conditioning (HVAC) systems has gained increasing attention. Despite the rapid development of machine learning techniques, the market is slowly adopting data-driven FDD as an alternative or complement to the traditional rule-based approaches. In this activity, data-driven FDD is defined as those that are trained or built from data using machine learning or multivariate statistical analysis methods [40] (rather than if/then logical rules and decision trees).

Similar to Activity C1, a literature review (Section 3.2) was firstly conducted to understand the state-of-the-art of data-driven FDD, including its general framework, reported methods for all processes within the framework, building components that data-driven FDD methods have been applied to, sources of data used in developing these FDD methods, how data-driven FDD methods are evaluated, and reported challenges for further development and market adoption of data-driven FDD.

Taking advantage of several other on-going activities of the participants, such as that at Lawrence Berkeley National Laboratory, this activity summarized an existing FDD data repository and software (Section 3.3). It is noted that not all FDD software reported in this activity is data-driven FDD. But this summary provides a comprehensive understanding of the market availability of FDD tools.

Based on the summarized literature, data repository, and existing FDD software tools, participants of this activity developed a roadmap (Section 3.4), aiming to guide industry stakeholders through the ecosystem of Automated Fault Detection and Diagnosis (AFDD).

3.2 Literature review summary

Based on the above definition of data-driven FDD, our examination of the literature covered the process of data-driven FDD, the systems studied, and the evaluation metrics employed. The data-driven FDD process encompasses several steps, including data collection, cleansing, preprocessing, baseline establishment, fault detection, diagnostics, and potential fault prognostics, as illustrated in Figure 5.

Several algorithmic methods were identified as being used in the FDD process, such as Clustering, Decision Tree (DT), Principal Component Analysis (PCA), Support Vector Machine (SVM), Support Vector Regression (SVR), Neural Networks (NN), Bayesian Networks (BN), Hidden Markov Models (HMM), Generative Adversarial Networks (GAN), and Ensemble Learning. While various data-driven methods have been investigated, there are few studies that compare the performance between methods in different categories (e.g., expert rule-based vs data-driven, supervised vs unsupervised).

Our review also found that data-driven FDD methods have been applied to various HVAC components and subsystems to detect and diagnose a range of faults. For large buildings, the focus has often been on Air Handling Unit – Variable Air Volume (AHU-VAV) systems, fan coil units (FCU), chillers, and boilers. The literature reported that 35% of the studies were dedicated to secondary AHU-VAV systems, with chillers following closely at 32% [38]. The AHU-VAV secondary systems, crucial for heating and cooling multiple zones, were often presented with actuator and equipment faults such as those in dampers, cooling/heating coil valves, fans, and air ducts [38].

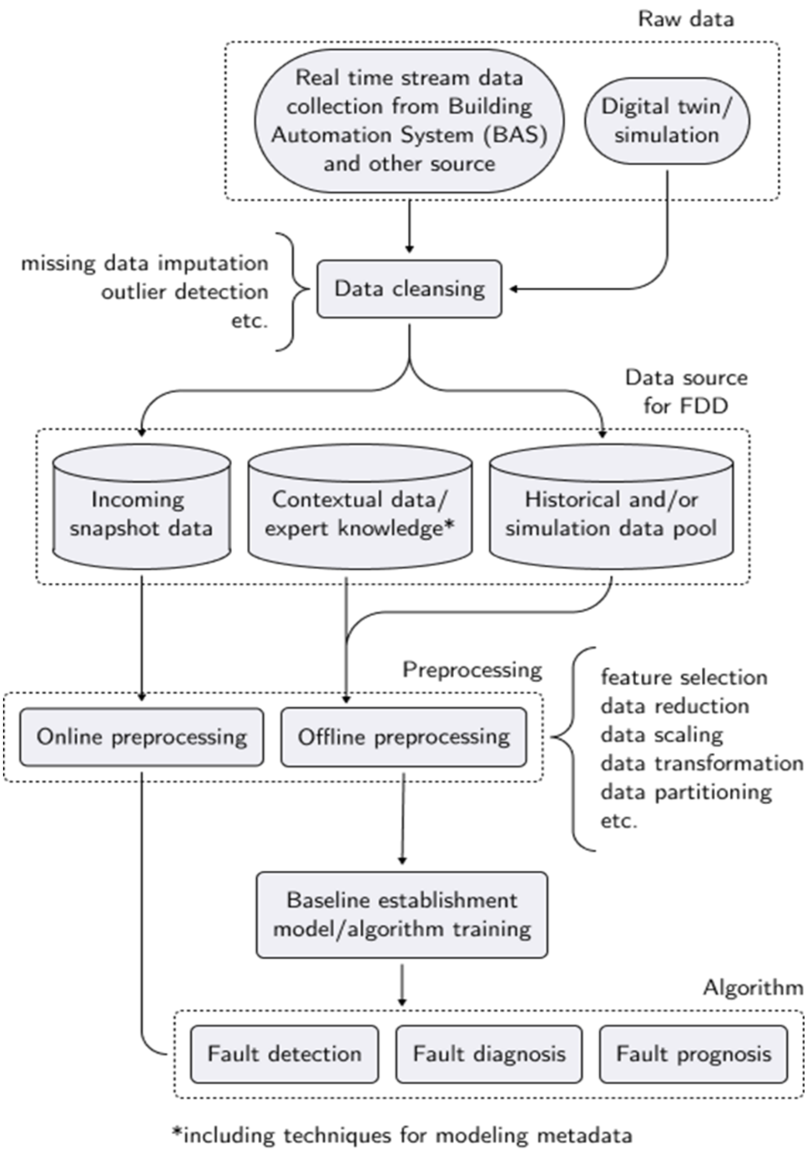

Figure 5 A General Data-driven FDD Process.

Chiller faults stand out as the most extensively studied in the context of data-driven methods in Vapor Compression Cycle (VCC) systems. These chiller faults are bifurcated into two categories: local faults, which include faults like condenser fouling, reduced condenser water flow, non-condensable in the refrigerant, and reduced evaporator water flow, and system faults such as, refrigerant leakage/ undercharge, refrigerant overcharge, and excess oil.

Additionally, a significant focus, accounting for 17% of the reviewed studies, has been directed towards whole-building level faults. The intricacies at this level arise from a confluence of factors such as building dynamics, external climatic conditions, system operating schedules and occupant comfort requirements. These collectively give rise to a myriad of building energy consumption patterns, which are not always straightforward to discern.

As for the sources of data used in developing these FDD methods, the literature revealed a mix of simulation data, laboratory experiments, and field measurements from real buildings. Among the papers reviewed, 48% used lab experiment data, 20% used simulation data, and 32% used real building data. The majority of whole building applications rely on real field measurement data while system-level VRF, AHU and Chiller applications mainly rely on laboratory data.

The evaluation of data-driven FDD is crucial, and the literature presents a gamut of dedicated metrics to achieve this. Broadly, these metrics fall into three categories:

- General Evaluation Metrics: These encompass fundamental measures like true positive rate (TPR), false negative rate (FNR), and correct diagnosis rate (CDR).
- Classification Problem Metrics: Tailored for data-driven classification problems, confusion matrix, accuracy of correct predictions, F-measure (or F-score), Receiver Operator Characteristic (ROC), and Area Under the Curve (AUC).
- Statistical Significance Tests: Useful for comparing different classification models in FDD, common tests include the t-test, McNemar's Test, and the Friedman Test.

Based on the findings from our review, we have identified some of the ongoing efforts and challenges to further the development and market adoption of data-driven FDD, as follows:

- Real-building deployment
- Performance Evaluation, Benchmarking, and Fault Impact Analysis
- Scalability and Transferability
- Interpretability
- Cyber Security and Data Privacy
- User Experience

Details on these challenges are presented in [38].

In summarizing this review, we aspire it to provide insights and directions for practitioners and researchers to develop the next generation data-driven FDD products. More details about the review are provided in [40].

3.3 Data repository and software summary

3.3.1 LBNL FDD data repository

In the past thirty years, the development of various FDD solutions for buildings has attracted significant attention around the world. However, a persistent challenge to ongoing development advances is a lack of common datasets and algorithm test methods, which are essential to support the vetting of new algorithms. In the past, very few publicly available, labeled FDD datasets have been published. Most research used research-project specific data, which were restricted by NDAs or other data sharing restrictions. Some early research produced a handful of FDD datasets, which are publicly available and have been widely used to develop FDD technologies. However, those datasets are limited to a few types of HVAC equipment. The fault types and faulty data range are very small. For example, the ASHRAE RP-1043 project included 8 fault types for chillers [41]. The ASHRAE RP-1312 project included 13 faults for air handling units (AHU) [42]. For both datasets, each fault type contains faulty data ranging from one day to a few days within one typical operational season.

To bridge this gap, researchers at Lawrence Berkeley National Laboratory developed the largest ever HVAC FDD dataset. It covers the most common HVAC systems and configurations in commercial buildings, across a range of climates, fault types, and fault severities. The time series points that are contained in the dataset include measurements that are commonly encountered in existing buildings as well as some that are less typical. Simulation tools, experimental test facilities, and in-situ field operation were used to generate the data.

The FDD dataset includes 7 HVAC systems, including the single duct AHU system, the packaged rooftop unit (RTU), the dual duct AHU system, the fan coil unit (FCU) system, the fan power unit (FPU), the boiler plant, and the chiller plant. Data for most systems spans faulty operation in one year. The total fault cases number 257 (i.e., faults at different severity levels), with an associated 8 billion data points.

The developed FDD datasets and associated inventories are fully open to the public with no cost, and can be downloaded from the website: https://faultdetection.lbl.gov/. Granderson et al. documented the development of the FDD data set [43, 44]. In addition, [45] presents a systematic framework for evaluating the performance of FDD algorithms.

3.3.2 FDD software tool repository

In the U.S., there is a thriving ecosystem of commercially available FDD software tools that are increasingly being adopted in the market. FDD software tools employ operational data collected from building BMS systems, sensors, and meters, to automatically detect equipment and control problems, or degrading performance in an HVAC system, and to diagnose potential root causes [46].

FDD software comes in different flavors. It can be hosted in cloud-based or on-premise servers external to the BMS, it can be run from desktop applications, or embedded in equipment [47]. In addition, building management systems often offer collections of rules that are packaged and sold as FDD libraries. FDD software that integrates with the BMS has been documented to save on average 9% in energy use, with two-year paybacks in portfolio implementations [46]. A few examples are provided in Table 1 - this list is not intended to be comprehensive[1]. Several HVAC FDD tools and corresponding vendor information are provided in Table 1.

Table 1 Examples of commercially available AFDD software tools

No	Company	FDD software name	Website
1	Clockworks Analytics	Clockworks	https://clockworksanalytics.com/
2	CopperTree Analytics	Kaizen	https://www.coppertreeanalytics.com/
3	Ezenics	Ezenics	https://ezenics.com/
4	Cimetrics	Analytika	https://cimetrics.com/
5	SkyFoundry	SkySpark	https://skyfoundry.com/
6	Prostar Energy Solutions	eIQ Platform	https://prostarenergy.com/
7	Iconics	Facility AnalytiX	https://iconics.com/
8	BuildingLogix	BuildingLogix Data Exchange	https://buildinglogix.net/
9	Lean FM Technologies	LEANFM RESCRIPTV	https://leanfmtech.com/
10	KODE Labs	KODE	https://kodelabs.com/
11	HITACHI	exiida	https://www.hitachi-gls.co.jp/products/exiida/monitoring/

[1] Reference herein to any specific commercial product, process, or service by its trade name, trademark, manufacturer, or otherwise, does not necessarily constitute or imply its endorsement, recommendation, or favoring by the authors, by the United States Government or any agency thereof, or by the Regents of the University of California.

12	azbil	BiG EYES	https://aa-industrial.azbil.com/ja/products/monitoring-control-system-software/monitoring-operation-support/big-eyes
13	CIM	PEAK	https://www.cim.io/
14	Bueno	Smart Building Analytics platform	https://www.buenosystems.com.au/
15	Switch BMS	Switch OpX - Operational Excellence	https://www.switchBMS.com/
16	Soundsensing	Soundsensing	https://www.soundsensing.no/
17	Ento	Ento AI Analytics	https://www.ento.ai/da/
18	Utilifeed	Fault Detection Utilifeed	https://www.utilifeed.com/
19	Climify	Human-Centered Building Monitoring and Control	https://climify.com/

3.4 Road map summary

This roadmap aims to guide industry stakeholders through the ecosystem of AFDD and facilitate increased adoption and deployment of AFDD in buildings and their systems - by addressing identified barriers and providing possible solutions for building owners, operators, facility managers, and technology providers. The entries to this AFDD roadmap originate from the latest review and survey studies on the topic [40, 48, 49], and inputs from the different participants of the activity C2.

Figure 6 illustrates the ecosystem of the implementation and operation of AFDD solutions in buildings. It includes: 1) AFDD development and implementation, 2) Running AFDD solutions in buildings, 3) Data management, and 4) Fault analysis and handling.

The mapping of the different barriers, stakeholders and potential solutions are categorized in Table 2 to Table 5 below, each according to one of the four sub-domains of the AFDD ecosystem defined in Figure 6. Each table divides the proposed solutions by each of five categories of identified barriers[2] that the proposed solution address. The identified barriers for the tables are as follows:

- **Economic and business:** Costs and benefits for end-users and/or business limitations.
- **Technological and technical:** Technical knowledge, interoperability, infrastructure and/or data.
- **User-related:** User experience, interface and/or misunderstanding.
- **Regulatory:** Policies, GDPR and/or cybersecurity.
- **Social and societal:** Cultural, community and stakeholders, benefits for society and/or environmental sustainability.

[2] Based on: Andersen et al. https://doi.org/10.1016/j.enbuild.2023.113801

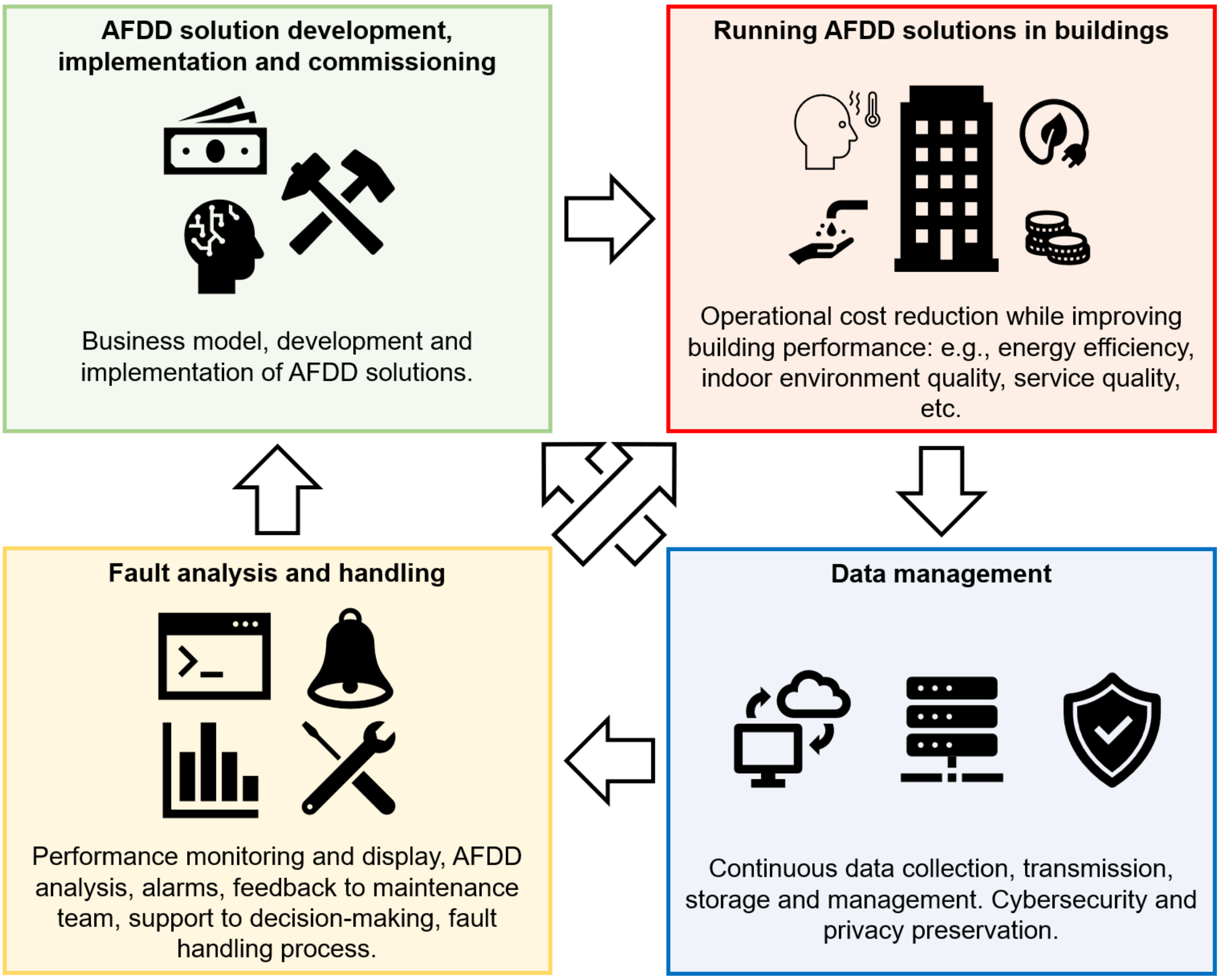

Figure 6 Mapping of the AFDD ecosystem: implementation and operation of AFDD solutions.

Also, in the Tables below, the term 'stakeholder' refers to a person or group with an interest in a topic that can affect or be affected by its operations and performance. The identified stakeholders are:

- **Building Technology industry (AFDD companies + BMS vendors).** The key issues affecting this category typically relate to the implementation of data driven AFDD in buildings, and are mostly related to *interoperability* and *market related issues and adoption*. Legacy BMS systems and proprietary communication protocols, in existing buildings, challenge the implementation of new AFDD products. This makes integration not only technically difficult but requires a significant investment that may not match the customer's expectations.
- **Building owners.** Such stakeholders might be required to make large investments to upgrade or retrofit their existing equipment. Implementing advanced AFDD tools may expose the building systems and data to cybersecurity issues (industrial clients may be the most concerned). Without a broadly accepted methodology to assess the potential performance of AFDD tools in operation, the calculation of KPIs like return on investment (ROI) or payback time may be troublesome and it could be difficult to estimate savings potential.
- **End users (maintenance staff).** Lack of interpretability or transparency behind the results of AFDD tools may lead to difficulty in accepting results (fault root causes, diagnosis, actions to be taken) from the end user perspective. They may be interested in fully understanding how the tool calculates a certain result or prioritizes the intervention on a certain fault. Along with trust issues and the learning curve required to understand and use such tools, the end users may be reluctant to change day-to-day operations in favor of new procedures.

Table 2 AFDD development and implementation

Barrier Category	Potential Barriers	Stakeholders	Existing Resources	New Possible Solutions
Economic and business	AFDD solutions can be too expensive, particularly when there are legacy systems leading to poor data access	AFDD companies, building owners	None identified	Information tools that help assess the return of investment on AFDD; and more clarity on AFDD costs
Technological and technical	Integration with underlying BMS is difficult due to lack of semantic information (lack of semantic interoperability)	BMS companies, AFDD companies, maintenance staff	There exists some but limited AFDD-related ontologies	Development and adoption of an ontology and semantic principles for the development of AFDD solutions
	Quality of BMS data is low: uncertain, missing data and a lack of labeled ground truth data	AFDD companies, building owners, BMS companies	Project Haystack and Xeto, ASHRAE 223p, Brick; semantic sufficiency; open source library online (but no centralized way to share)	Foster more open-source tools, datasets, models and benchmarking; FAIR principles; development of guidelines to indicate how much sensing is required.
	Scalability and transferability (model), portability (service); data schema; similarity	BMS companies, AFDD companies, researchers	ASHRAE 223p, Brick; semantic sufficiency; open source library online (but no centralized way to share)	Use an ontology; open source portable applications; transfer learning; measure the accuracy when a building model is transferred
User-related	Owner does not understand the economic/operational benefits	Building owners, AFDD companies	Studies on energy and cost reduction with AFDD; DOE report	Assessment of the return of investment on AFDD; training people to understand the value of data
	Industry generally has low awareness of AFDD (conservative for innovations due to uncertainties)	BMS companies, building owners, maintenance staff	Research on AFDD benefits; ROI analysis	Demonstrators; training people to understand the value of data
Regulatory	Lack of standardization; data privacy and security; cost implications; technical; complexity and expertise	BMS companies, AFDD companies, researchers	Incentives and policies; building codes and regulations	Adoption of smart readiness indicator in EU

Social and societal	Public resistance due to perceived risks; impact on employment in traditional maintenance roles; societal undervaluing of long-term sustainability.	General public, labor unions, educational institutions	Public forums and discussions; social studies on technology adoption	Public engagement initiatives; collaboration with educational institutions to integrate AFDD technologies into curricula; campaigns that highlight environmental and economic benefits.

Table 3 Running AFDD solutions in Building.

Barrier Category	Potential Barriers	Stakeholders	Existing Resources	New Possible Solutions
Economic and business	Customers are reluctant to pay regular service fees for AFDD and cloud-based solutions, they prefer a one-time fee.	Maintenance staff, building owners	None identified	Development of demonstrator projects to showcase benefits; detailed business cases illustrating long-term savings; transparency in pricing and the breakdown of service fees; payment systems offering various levels of service.
Technological and technical	Integration with existing systems; handling large volumes of data; ensuring accuracy and relevance of alarms.	BMS companies, AFDD companies	None identified	Enhanced interoperability standards; development of advanced predictive algorithms to reduce false positives; improved user interfaces for easier interaction with AFDD systems; cloud solutions with scalable data management.
User-related	Cooperation with maintenance staff	BMS companies, AFDD companies, maintenance staff	Commissioning training processes; existing educational materials from tool developers.	More focused, hands-on training sessions; simplification of AFDD interfaces; regular update workshops; creation of easy-to-follow maintenance protocols; incentivizing staff through certification in AFDD technology.
	How to better guide the staff to do maintenance, how to make them understand FDD tool outputs (e.g., too many alarms)	AFDD companies, maintenance staff	Training (commissioning training process, tool developer to educate)	Focused training of maintenance staff on AFDD solutions provided by AFDD companies on specific BMS; develop better UI to support decision making to correct faults

Regulatory	Compliance with data regulations; industry standards for AFDD implementation.	AFDD companies, maintenance staff, government regulatory bodies, industry standard organizations, legal departments.	General data protection regulations; existing building and industry standards.	Engagement with regulatory bodies to develop AFDD-specific standards; workshops and seminars on regulatory compliance for AFDD vendors and users; development of compliance toolkits tailored to AFDD.
Social and societal	Public skepticism about BMS in building management; concerns over data privacy; impact on employment for traditional roles.	Building owners, General public	None identified.	Public awareness campaigns about the benefits of AFDD; engagement with privacy advocates to ensure robust data protection; training programs aimed at upskilling traditional roles into AFDD-related positions; community outreach programs to educate on sustainability benefits of AFDD.

Table 4 Data Management.

Barrier Category	Potential Barriers	Stakeholders	Existing Resources	New Possible Solutions
Economic and business	High costs associated with data storage and processing; economic feasibility of scaling data infrastructure; ROI uncertainty.	Building owners, AFDD companies, investors, IT service providers.	Limited financial analysis on data management for AFDD; existing data storage solutions.	Cost-benefit analysis of data infrastructure; development of scalable cloud storage solutions to reduce upfront costs; subscription-based data management services; financial models highlighting long-term savings from efficient data management.
Technological and technical	Many sensors have silo-based data extraction. Might need additional programming and knowledge for a continuous data stream.	BMS companies, AFDD companies, maintenance staff	From a software perspective, this is easy. However, the associated cost might be a barrier. In particular, indoor environmental sensors sometimes have individual cloud services. Having multiple cloud services is difficult to manage costs.	Development of universal data adapters or middleware to integrate disparate systems; investment in training for IT staff on data integration; industry-wide standards for sensor data outputs; cost/benefit study to justify integration investments.

User-related	Lack of technical expertise in handling and interpreting large datasets; resistance to new systems due to complexity.	Facility managers, maintenance staff, building owners, AFDD companies.	Basic training materials; some AFDD interfaces designed for non-expert use.	Comprehensive training programs on data management; development of more intuitive, user-friendly AFDD dashboards; ongoing support and troubleshooting services; community forums for sharing best practices and troubleshooting.
Regulatory	GDPR / privacy for data sharing can be a limitation to access and use data for AFDD, especially if the costumer does not see the benefits of sharing their data.	BMS companies, AFDD companies	General understanding and compliance frameworks for GDPR.	Detailed explanations and transparent policies on data usage; secure data handling and storage practices; development of consent protocols that clearly benefit users; workshops and seminars on the benefits of data sharing under regulatory compliance.
Social and societal	Public concerns over surveillance and data misuse; social resistance to perceived over-reliance on technology.	General public	Public forums; existing laws and regulations on data privacy.	Public engagement initiatives to discuss and address data privacy concerns; collaborations with privacy companies to build trust; educational campaigns highlighting the environmental and operational benefits of AFDD; robust anonymization techniques to protect individual privacy while using data for AFDD.

Table 5 Fault analysis and handling.

Barrier Category	Potential Barriers	Stakeholders	Existing Resources	New Possible Solutions
Economic and business	Unclear business models for fault handling and analysis.	Maintenance team, building owner	Engaging management to reduce energy use and improve indoor environment in buildings.	Development of clear business models outlining ROI for fault analysis; performance-based contracting; incentive models for energy savings achieved through effective fault management.
Technological and technical	Each building is different; diversities of control loops; how to measure similarities; how to compare performance of buildings	AFDD companies, maintenance team, building owner	LBNL benchmarking data; simulation	New KPIs; need standard way to organize data (e.g., Energy Star); compare energy performance with NABERS energy rating

	Difficulty of creating baselines, label data; taxonomy of the fault; method to identify severity of the fault is not mature; difference between simulation (lack detail) and real-world; how to measure the impact	Detail modeling (Modelica); ontology-developers; researchers	ASHRAE 223p Brick	
	Interpretability; black box; deep learning structure; system complexity (even rule-based is difficult)		LIME, SHAP; LLM, generative AI	Explainable AI; hybrid model that combines both physics and data-driven approach.
User-related	Maintenance staff try to hide their faults	AFDD companies, maintenance team, building owner	Collecting the control signals, sequences	Enhanced training programs focused on the benefits and operation of AFDD tools; development of fault prioritization systems to manage alarm frequency; user-friendly interfaces and dashboard customizations to aid maintenance staff in managing and responding to alerts more effectively.
	How to understand FDD tools to better support the maintenance staff (e.g., too many alarms), how to better guide the staff to do maintenance	AFDD companies, maintenance team, building owner	Basic training of BMS, control systems and sequences.	
Regulatory	Lack of specific regulations or standards on AFDD in buildings.	Regulatory bodies, AFDD companies, building owners.	General building and safety regulations.	Creation of specific standards and regulations for AFDD systems; integration of AFDD requirements into existing building codes and standards; workshops and seminars to engage regulators and stakeholders in establishing these regulations.
Social and societal	Public skepticism towards automated systems and potential job displacement concerns.	General public, educational institutions.	Public forums and debates on BMS and employment.	Public awareness campaigns that highlight the environmental and economic benefits of AFDD; educational programs that provide training for new technological roles

4. Activity C3: Building-to-Grid Applications

4.1 Introduction

To address the environmental and sustainability challenges in the energy sector, and to enable the transition from fossil fuels to fluctuating renewable energy sources (RES), a drastic change is required in the operation of current energy systems, energy grids and in the buildings that are connected to them (see Figure 7).

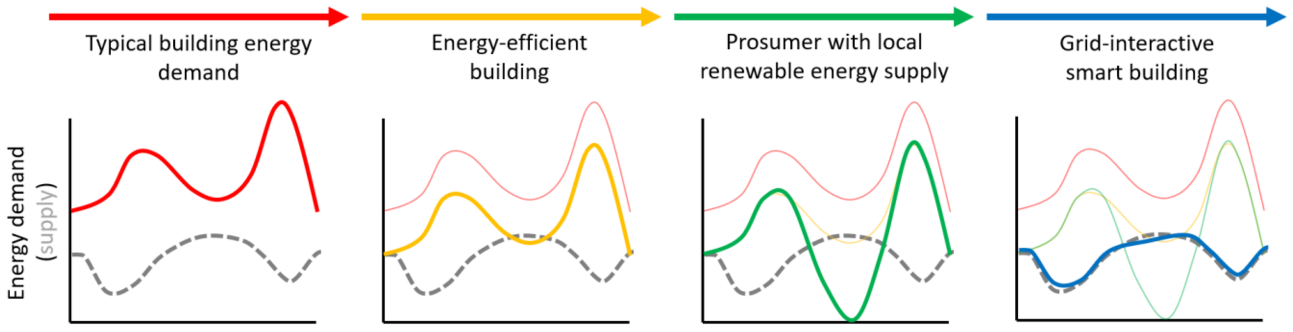

Figure 7 Paradigm shift: Toward building-to-grid services from energy-efficient smart buildings connected to sustainable grids dominated renewable energy sources [50].

Buildings have a certain energy flexibility potential. They have the ability to adapt or modulate their short-term (a few hours or a couple of days) energy demand and energy generation profile according to climate conditions, user needs and energy network requirements without jeopardizing the technical capabilities of the building systems and the comfort of occupants.

Building energy flexibility strategies (in the form of demand response) enable load control/modulation to provide building-to-grid (B2G) services to the local energy grids. These B2G services support the matching of the energy demand profile with the energy supply profile in smart grids dominated by RES. They also help to tackle other grid challenges such as voltage and frequency stability in electrical grids, peak power limitations and local bottleneck effects in thermal and electrical grids, high costs and CO_2-intensive operation of peak power generators, negative electricity prices, costly reinforcement or extension of the networks, or accelerated deterioration of hydronic networks caused by the unstable operation of thermal grids.

Currently, most of the large-scale orchestration of B2G services, provided by a multitude of different decentralized prosumers and energy flexibility assets, is driven by incentive signals broadcast by energy grids (e.g., dynamic energy price or grid CO_2 intensity). Energy end-users providing B2G services can thus be rewarded by optimizing building operation to minimize operational costs from the dynamic energy price on a short time horizon (typically 25 hours or up to a few days). As an alternative, the B2G service provider can be directly paid for completing a specific demand response action, e.g., reducing peak power demand by a pre-agreed amount over a specific time window.

The development, testing, analysis and optimal management of large numbers of decentralized flexibility assets and numerous B2G service providers, requires a robust streamlined and interoperable demand-response assessment framework comprising KPIs covering the different scenarios, system configurations and stakeholders. It should also include reliable energy demand baseline models to establish the counterfactual reference energy profile of a building when performing demand response [39, 51].

Activity C3 focused on developing an online platform to gather, evaluate, compare, present and promote concrete cases of buildings providing services (such as demand response services) to the energy grid.

To that end, this activity addresses some important research gaps such as context, definitions and KPIs for building energy flexibility assessment, and the heterogeneity of data representations in datasets of B2G services. It promotes the use of ontologies and semantic principles to standardize the definition and computation of KPIs. It provides a structure for a Python package to streamline B2G service assessment from commonly available operational building data.

A semantic/ontology-based approach appears to be the best way to ensure the interoperability of the different elements in the B2G ecosystem, and to support the portability of B2G services across heterogeneous buildings. In this way building energy flexibility assets and demand response resources can become digitally enabled and increasingly 'smart' [52, 53].

The assessment tools of this Python package are implemented in an online web-application show-casing B2G service cases. It enables any user to rapidly perform data-driven energy flexibility analysis on open-access or uploaded datasets.

The Activity C3 workflow - from research to target objective – is illustrated in (Figure 8). It is believed that the results, framework, recommendations and tools developed within the activity C3 will contribute to foster the large-scale development of B2G services for all types of energy networks in future sustainable smart energy grids dominated by renewable energy sources [51].

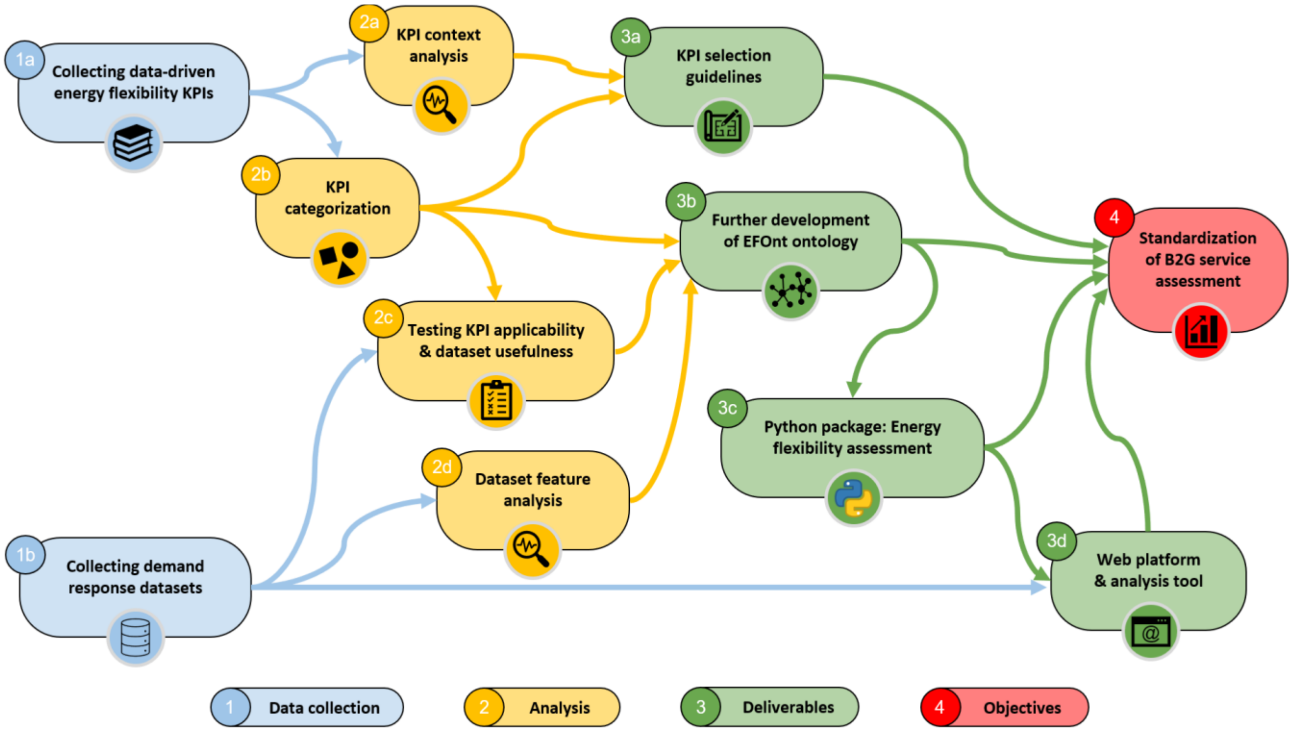

Figure 8 Workflow of the IEA EBC Annex 81-C3 activity on B2G applications and services [51].

4.2 Literature review summary on data-driven energy flexibility KPIs

As discussed above, an increasing number of studies are quantifying the energy flexibility and demand response of single buildings and clusters of buildings. However, most of these studies rely on numerical simulations and perform energy flexibility assessments that are not necessarily possible for real buildings. In that context, activity C3 conducted a literature review of data-driven KPIs that are suitable for the operational phase of buildings performing B2G services and demand response.

It was found that most studies focus on residential buildings (48.9%) and commercial buildings (28.3%). The application scale is mainly at the single-building level (53.3%) and building cluster level (41.3%). Only 25.8% of the studies involve real measurements, 65.2% rely on numerical simulations [39].

The review highlights that the two main constraints in quantifying energy flexibility through operational building data analysis are (i) the lack of robust data-driven approaches for generating baseline load profiles without demand response activation (which are necessary for calculating baseline-dependent KPIs) and (ii) the lack of KPIs that can be computed without need of baseline or reference scenario inputs (i.e., baseline-free KPIs).

A total of 81 distinct data-driven KPIs were identified in the reviewed scientific literature on building demand response. These KPIs can be classified into 12 core energy flexibility categories:

- Peak power shedding
- Energy/average power load shedding
- Peak power/energy rebound
- Valley filling
- Load shifting
- Demand profile reshaping
- Energy storage capability
- Demand response energy efficiency
- Demand response costs/savings
- Demand response emission/environmental impact
- Grid interaction
- Impact on indoor environmental quality

In addition, 29 other KPIs that are not directly related to energy flexibility are typically found in studies about demand response. These KPIs belong to 4 more generic building performance categories:

- Energy efficiency
- Costs and savings
- CO2 emissions/environmental impact
- Grid interaction

Table 6 and Table 7 below show the most popular (most frequently used in scientific studies) data-driven energy flexibility KPIs.

Table 6 Most popular baseline-required KPIs for assessing demand response and energy flexibility [39, 51].

KPI denomination	Definition
Energy efficiency of demand response action	The difference in total energy use between the scenario with demand response and the reference scenario without demand response divided by the difference in energy use during the length of the demand response action between the scenario with demand response and the reference scenario without demand response
Flexibility savings index	The ratio between the energy costs of the scenario with demand response and the energy costs of the reference scenario without demand response
Peak power shedding	The difference between the peak power use of reference scenario without demand response and the peak power use of the scenario with demand response

Table 7 Most popular baseline-free KPIs for assessing demand response and energy flexibility [39, 51].

KPI denomination	Definition
Flexibility factor	The difference between the energy use during non-peak periods and peak periods divided by the sum of energy use during non-peak periods and peak periods
Energy shift flexibility factor	The difference between the energy use during low-price periods and high-price periods divided by the sum of energy use during low-price periods and high-price periods
Load factor	The ratio between the average power use and the maximum power use

81% of the data-driven energy flexibility KPIs found in the scientific literature require a baseline to be computed. As mentioned above, establishing a baseline energy profile (a counterfactual energy demand when no demand response event occurs) is challenging. Ideally, a baseline estimation should be robust, transparent and prevent the possibility of manipulating B2G service market with reward mechanisms based on baseline-required KPIs. However, there is currently no consensus about which data-driven energy demand baseline generation method would perform best, especially at low aggregation levels [39, 51]. The most commonly used data-driven methods that are applicable for single-building and district energy demand baseline generation are as follows:

- **Control group methods**: Construct a baseline from monitoring data of buildings that are similar to the target ones with equivalent boundary conditions (weather, occupancy, operation) but do not perform any demand response at the time of evaluation [39].
- **The averaging methods** (similar day look-up approach or *XofY*): One of the most popular *XofY* load estimation technique is the *HighXofY*, which takes the average load of the *X* highest demand days from a set of *Y* admissible days prior to the demand response event [39].
- **Regression models**: Load forecasting is often performed with robust autoregressive models, such as ARMA (Auto Regressive Moving Average), ARIMA (Auto-Regressive Integrated Moving Average), GAM (Generalized Additive Model), or LASSO (Least Absolute Shrinkage and Selection Operator). However, these models may require large historical data to be fitted correctly [39].
- **Shallow machine learning methods**: Currently, many popular machine learning methods employ relatively simple models with a small number of layers or processing stages (shallow artificial neural networks, decision trees, random forests). These models present a limited capacity to learn complex and non-linear patterns from multi data with high dimensionality. They are thus only adequate for data with relatively simple patterns and straightforward relationships between features and outputs.
- **Deep machine learning methods**: In recent years, deep machine learning methods have emerged to leverage deep neural networks (DNN) with a very large number of hidden layers and neurons, recurrent architecture and attention mechanisms. These DNNs are very well suited to learn intricate patterns and representations from time series data (typical dynamic data from building systems) and generate forecasts of building energy profile and indoor environment variations (sequence-to-sequence forecasting) [54, 55]. In particular, long short-term memory and time-delay neural networks, have gained popularity for building energy profile forecasting. However, DNNs require a large amount of training data to outperform more simple and robust statistical methods. Large building operation datasets with sufficient quality for DNN training are scarce, but this limitation can be mitigated by employing transfer learning principles and synthetic data generators [56].
- **Hybrid models**: Combining some of the abovementioned modeling approaches has also been explored to perform load demand forecasting.

4.3 Building demand response dataset collection and analysis

In order to develop, study, test and benchmark B2G services, at scale, one must have access to diverse datasets from buildings and clusters of buildings performing demand response. Although several studies have produced such simulation or monitoring data to test their hypothesis, getting access to it is very difficult. Activity C3 attempted to collect all open-access building energy flexibility datasets (in the form of time series data, along with the appropriate metadata and case description) from publicly available datasets and data platforms (e.g., Kaggle, Data-in-Brief), scientific publications, and via direct contact with research teams and scientific communities (e.g., IEA EBC Annex 81, 82, 83, and 84).

However, from the 330 datasets identified as potentially of interest, only 16 were actually deemed adequate with proper descriptions and open access (or soon to be open access) data (see Figure 9). A very large share of the dataset candidates were miscategorized, out of scope, without sufficient description or unavailable to participants outside of the original research group that had generated the data [39]. This denotes a clear lack of open dataset culture in the building demand response community, which can hinder future developments in this field. It is thus highly recommended to put more effort into curating, describing and sharing future datasets generated in upcoming B2G service studies and pilot projects.

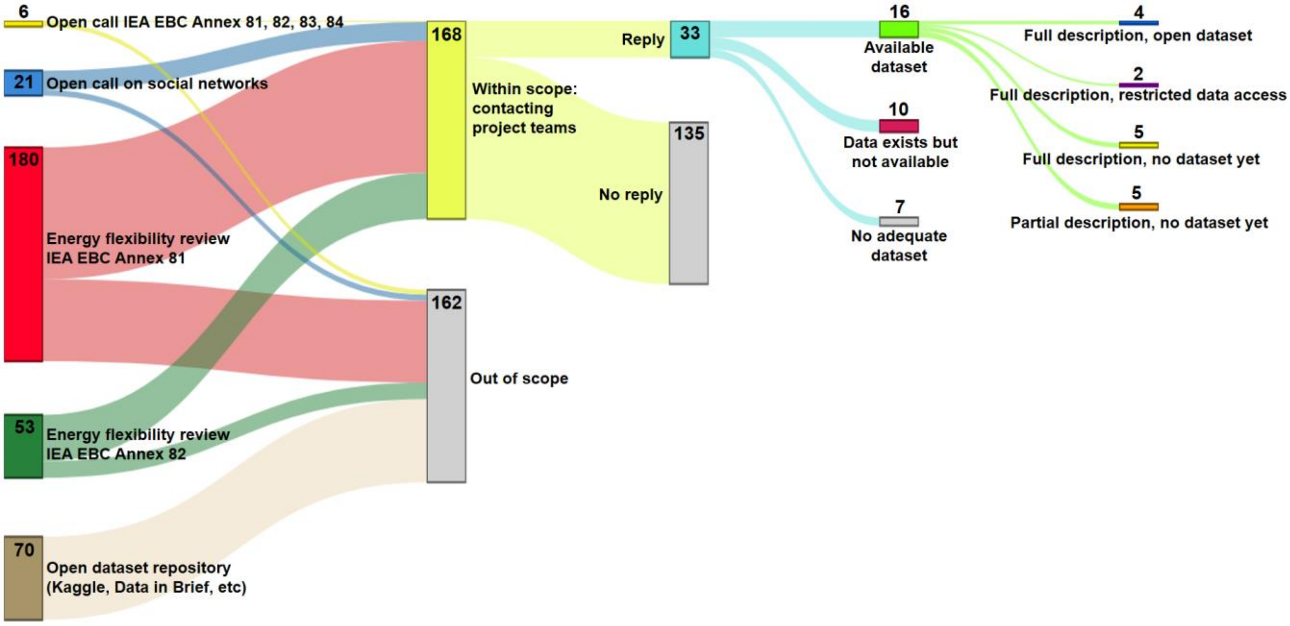

Figure 9 Building demand response dataset collection campaign by the IEA EBC Annex 81-C3 activity [39].

The 16 collected datasets represent a wide variety of building energy flexibility studies. They included data from real monitored buildings, hardware-in-the-loop setups, and numerical simulations with different building typologies. Most datasets were associated with flexibility in electrical grids, and only a few were connected to district heating networks.

Most of the reported demand response schemes were based on time-of-use, real-time pricing, and flat-rating pricing tariff programs. Load shifting and load shedding were the most common flexibility modes. HVAC systems were the most frequent resources to deliver flexibility, often triggered by temperature adjustments.

By comparing the primitive variables required for calculating the different data-driven building energy flexibility KPIs and the features of the collected B2G datasets, one can see in Table 8 that there is not a very good match between the former and the later. This suggests that the collected datasets have limited usefulness for performing demand response assessment, and that most of the reviewed KPIs are restricted in their applicability.

The most commonly available features across datasets were indoor temperature, followed by end-use energy demand, thermostat setpoints, occupancy and price signal. Since event timing and power demand are among the most critical variables required by the KPIs, and most datasets do not include them, additional modelling and calculations are needed to derive them. For instance, the price or emission signal can be used to define event timing and energy consumption to determine power demand.

Overall, based on the KPI data requirements and data availability, the three most easily calculated energy flexibility KPIs are demand response energy efficiency, demand profile reshaping and energy/average power load shedding. One can also note that the value of a dataset for KPI computation does not increase with the number of variables it contains. While some datasets have many variables, they may not have the most commonly required ones for demand response assessment [39, 51].

Table 8 Input variables required by the KPIs vs available ones in the collected B2G datasets [39, 51]

Primitive variables	**% required by KPIs**	**% available in datasets**
Event timing	37.66%	18.75%
Energy consumption	35.06%	81.25%
Power demand	32.47%	6.25%
Event request action	24.68%	37.50%
Price signal	16.88%	50.00%
Energy generation	12.99%	25.00%
Event request size	11.69%	0.00%
Indoor temperature	5.19%	93.75%
Thermostat setpoint	5.19%	62.50%
Emission signal	3.90%	12.50%
Storage volume	2.60%	0.00%
Monetary incentives	2.60%	0.00%
Occupancy	1.30%	56.25%
Indoor CO2	1.30%	12.50%

4.4 Toolbox for data-driven assessment of energy flexibility

At the moment, there is no standard way to assess B2G services. The different building stakeholders, building owners, tenants, building managers, policymakers, utility companies, and grid operators employ various KPIs to evaluate the effectiveness of flexibility assets and to ascertain the viability of new technologies, policies, demand response programs and control strategies.

To address this limitation, Activity C3 has developed an open-source toolbox in the form of a Python package (*energy-flexibility-kpis*) for the data-driven assessment of demand response and energy flexibility of buildings. This Python package leverages the EFOnt ontology [57] to apply semantic principles for the standardization of KPI definitions and computation. A semantic/ontology-based approach appears to be the best way to streamline demand response assessment from commonly available operational building data. It ensures the interoperability of this toolbox with the different elements in the B2G ecosystem, and supports the portability of the B2G services across heterogeneous buildings.

Combined with other relevant ontologies, representing various useful knowledge domains for B2G services, the EFOnt ontology enables the creation of semantic data models that can facilitate the standardization of the demand response KPIs' definition, their data specification/requirements, data collection procedure, pre-processing, computation procedure and visualizations (see Figure 10). The demand response and building energy flexibility KPIs found in the scientific literature are progressively implemented in the *energy-flexibility-kpis* Python package, together with all necessary data treatment sub-functions and key data-driven methods for generating an energy profile baseline [51, 57] .

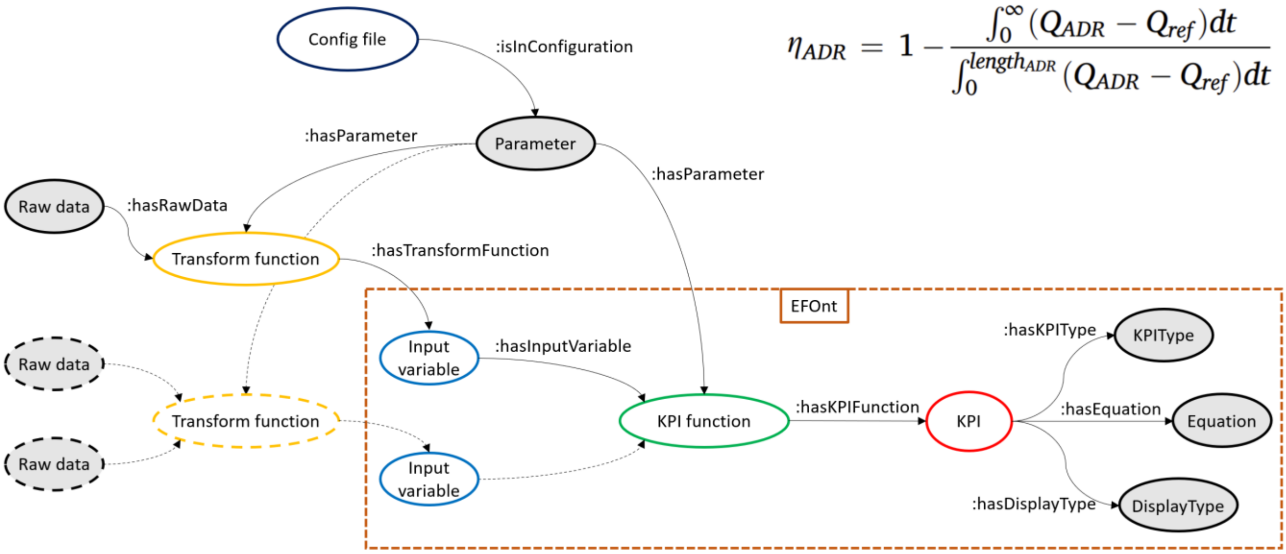

Figure 10 Semantic description of an energy flexibility KPI and its input variables in the EFOnt ontology. The different properties and attributes of the KPI (i.e., the denomination of the KPI, its associated equation, its type, the way this KPI should be displayed, its associated Python function, the list of needed input variables for the computation of the KPI) are defined within the EFOnt ontology. The EFOnt ontology is extended to integrate the definitions of the transform functions that are needed to convert the raw data from the building operation into the required input variables to compute the KPI, together with the parameters contained in a configuration file that indicate the settings of the assessment (e.g., evaluation window, or resampling methods) [51, 57].

This *energy-flexibility-kpis* Python package for the assessment of demand response and energy flexibility strategies can be found in the dedicated GitHub repository https://github.com/HichamJohra/energy_flexibility_kpis (under development) and can be installed from pypi.org (https://pypi.org/project/energy-flexibility-kpis/): *pip install energy-flexibility-kpis*.

4.5 Online platform for collection and analysis of B2G datasets

To showcase B2G services and the ease of use of the building energy flexibility assessment tools presented above, the *energy-flexibility-kpis* Python package was implemented into a data analysis workflow on an online web-app (see Figure 11). This online platform also includes an ontology explorer to navigate through the EFOnt ontology and the latest implementation of the *energy-flexibility-kpis* Python package together with a comprehensive selection interface to identify the most adequate demand response KPIs, depending on the type of stakeholder, data availability/features, performance goals and flexibility assets.

Moreover, the collected open-access datasets of buildings performing demand response are listed and loaded onto that web-app as examples. One can thus easily select those B2G service cases and perform energy flexibility assessment and comparison with all KPIs implemented in the Python package. In addition to showcasing various examples of B2G services, this online platform encourages users to share their datasets in open access.

This online platform (web-app) for the collection and analysis of building demand response datasets can be accessed with the following link: https://aau-ef-kpi-web-app.build.aau.dk/

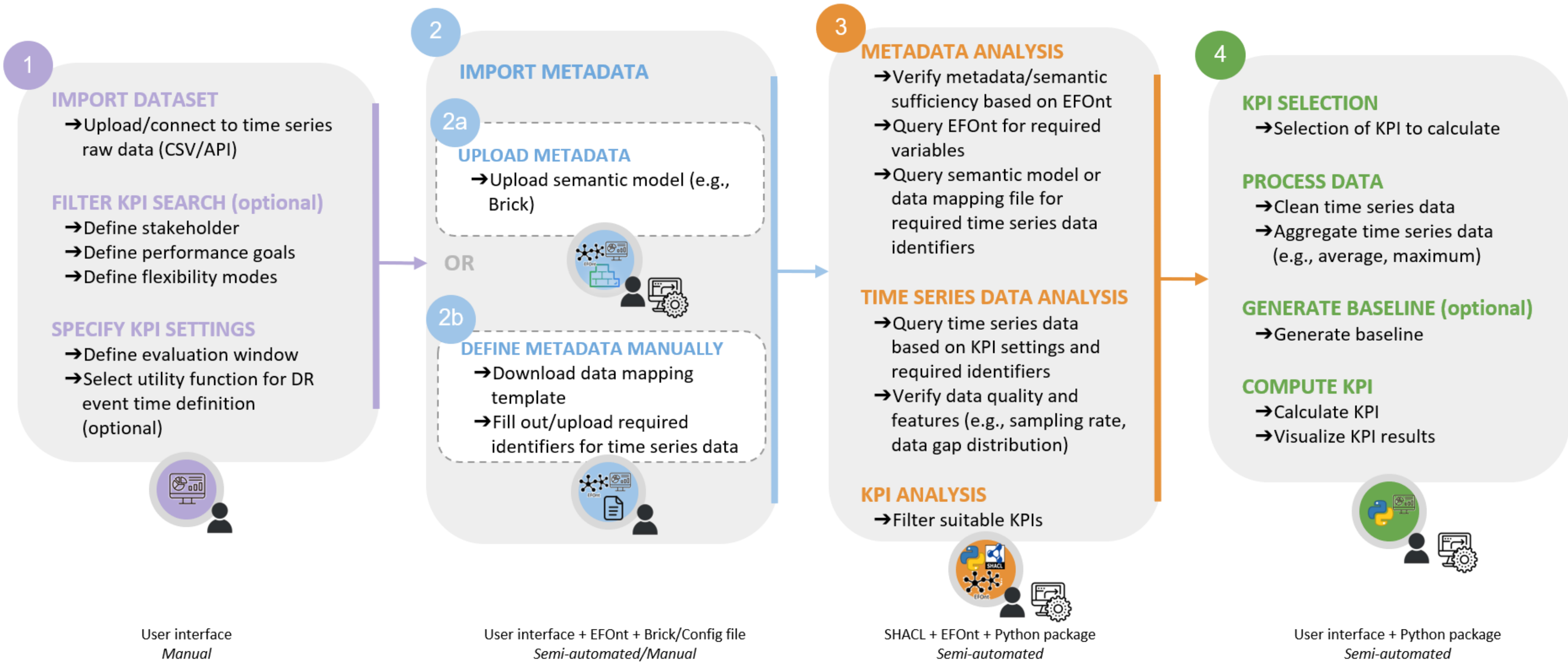

Figure 11 Data-driven energy flexibility quantification process for B2G applications [51].

5. Conclusions and future direction

During the course of ANNEX 81, Subtask C has engaged experts from many countries across four continents. Findings from Subtask C have led to the publication of three journal papers and one conference paper.

We (i) performed case studies to evaluate the feasibility of implementing reported building performance KPIs using data from five real buildings, (ii) collected 16 datasets for building energy flexibility studies, and (iii) assisted the development of a comprehensive data repository for FDD studies. An online platform with open-source toolbox was developed for analysis of B2G services and building energy flexibility evaluation. More specifically,

Activity C1 examined the KPIs of building operational performance with three key focuses:

- What KPIs exist in the building and energy-related literature,
- what KPIs can be implemented in existing infrastructures of buildings, and
- what KPIs are important to stakeholders.

Over 400 KPIs were identified, collected and reviewed, spanning building performance areas related to occupants [4], building operation, grid [39], and smart technology. Despite the abundance of proposed KPIs in the literature, challenges persist due to unclear definitions, unspecified sensor/meter data requirements, and a lack of real-life contextualization, especially at the whole-building level. The case study revealed that only about one-quarter of the gathered KPIs could be computed using BMS data from five case study buildings, with thermal comfort and energy-related KPIs being the most readily calculable.

The case study's results align with literature findings, underscoring that while these KPIs can be computed, their definitions often overlook the complexity of real-world building scenarios, introducing ambiguity and compromising reliability in calculations. Moreover, the stakeholders survey indicated that the KPIs related to occupants' health are the most important among all others. Nevertheless, it is worth noting the substantial divergence in the priority of building performance aspects among individual stakeholders.

In conclusion, there is a clear misalignment among KPI definitions in the literature, the data that is readily available from the current BMS, and stakeholders' needs. Further research is needed to contextualize KPIs across diverse application scenarios while considering stakeholders' perspectives. This step is essential for bridging existing gaps and ensuring a more cohesive integration of KPIs into the intricate landscape of building operational performance benchmarking.

Activity C2 defined the concept of data-driven FDD algorithms as those that are trained or built from data using machine learning or multivariate statistical analysis methods. Based on this definition, a thorough literature review was performed to understand the following topics related to data-driven FDD:

- What are typic processes and methods of data-driven FDD?
- What kind of building systems have data-driven FDD been applied to?
- What are the reported faults for these systems?
- How to evaluate data-driven FDD?

The literature review [38] revealed that many promising methods and frameworks are reported for implementing data-driven FDD, step by step from collecting data to detecting anomalies, to isolating root causes. However, there is a lack of data-driven fault prognosis studies.

Data-driven FDD has been applied to most building systems. Nevertheless, the majority of the reported studies are based on simulated or laboratory experimental data. There are comprehensive evaluation metrics developed for data-driven FDD performance evaluations.

To further promote market adoption of data-driven FDD methods, real time and real-building implementations are needed. Consequently, Activity C2 compiled a repository of FDD data and software, leveraging various on-going activities of the participants. Based on findings from the literature review and data repository, a roadmap was developed attempting to guide industry stakeholders through the ecosystem of FDD. The ecosystem includes 1) FDD development and implementation, 2) Running AFDD solutions in buildings, 3) Data management, and 4) Fault analysis and handling.

Five categories of barriers were identified for each part of the above ecosystems. These were economic and business barriers; technological and technical barriers; user-related barriers; regulatory barriers; and social and societal barriers. Stakeholders and potential solutions for each barrier in each part of the FDD ecosystem were identified and discussed in the roadmap. The roadmap is expected to facilitate deeper market adoption and deployment of data-driven FDD.

Activity C3 aimed to contribute to the standardization of B2G service assessment. Collaborating with C1, Activity C3 firstly identified 81 data-driven KPIs that are essential for the operational phase of building energy flexibility assessment. These KPIs can be classified into 12 core energy flexibility categories.

It was found that most studies relied on numerical simulations (65.2%) [39]. The review highlighted two main constraints in quantifying energy flexibility through operational building data analysis. These are (i) the lack of robust data-driven approaches for generating baseline load profiles without demand response activation (which are necessary for calculating baseline-dependent KPIs) and (ii) the lack of KPIs that can be computed without baseline or reference scenario inputs (baseline-free KPIs).

To help develop and benchmark B2G services at scale, Activity C3 collected 16 datasets that represent a wide variety of B2G studies and include data representing various building types from real buildings, hardware-in-the-loop testbeds, and numerical simulations. It was found that there is a poor match between the primitive variables required for calculating the different data-driven building energy flexibility KPIs and the features of the collected B2G datasets. This denotes a limited usefulness of the collected datasets for performing demand response assessment, and a restricted applicability of most reviewed KPIs.

The three most easily calculated energy flexibility KPIs are demand response energy efficiency, demand profile reshaping and energy/average power load shedding. Leveraging the EFOnt ontology [57], Activity C3 developed an open-source toolbox in the form of a Python package (*energy-flexibility-kpis*) for data-driven assessment of demand response and energy flexibility in buildings. The assessment tools of this Python package are implemented in an online web-application show-casing B2G service cases. They enable any user to rapidly perform data-driven energy flexibility analysis on open-access or uploaded datasets.

It is believed that the results, framework, recommendations and tools developed within the activity C3 will contribute to foster the large-scale development of B2G services for all types of energy networks in future sustainable smart energy grids dominated by renewable energy sources [51].

References

[1] J. C. Van Gorp, "Using key performance indicators to manage energy costs," *Strategic planning for energy and the environment,* vol. 25, no. 2, pp. 9-25, 2005.

[2] H. Li, Z. Wang, and T. Hong, "Occupant-Centric key performance indicators to inform building design and operations," *Journal of Building Performance Simulation,* vol. 14, no. 6, pp. 814-842, 2021.

[3] S. Verbeke, D. Aerts, G. Reynders, Y. Ma, and P. Waide, "Final report on the technical support to the development of a smart readiness indicator for buildings," *European Commission: Brussels, Belgium,* 2020.

[4] S. Sleiman *et al.*, "Overview of occupant-centric KPIs for building performance and their value to various building stakeholders," *Energy and Buildings,* p. 114704, 2024.

[5] H. Van Der Veer and A. Wiles, "Achieving technical interoperability," *European telecommunications standards institute,* 2008.

[6] M. Ginocchi, A. Ahmadifar, F. Ponci, and A. Monti, "Application of a smart grid interoperability testing methodology in a real-time hardware-in-the-loop testing environment," *Energies,* vol. 13, no. 7, p. 1648, 2020.

[7] T. Ford, J. Colombi, S. Graham, and D. Jacques, "The interoperability score," in *Proceedings of the Fifth Annual Conference on Systems Engineering Research*, 2007, pp. 1-10.

[8] M. van Amelsvoort, C. Delfs, and M. Uslar, "Application of the interoperability score in the smart grid domain," in *2015 IEEE 13th International Conference on Industrial Informatics (INDIN)*, 2015: IEEE, pp. 442-447.

[9] G. Pinto, R. Messina, H. Li, T. Hong, M. S. Piscitelli, and A. Capozzoli, "Sharing is caring: An extensive analysis of parameter-based transfer learning for the prediction of building thermal dynamics," *Energy and Buildings,* vol. 276, p. 112530, 2022.

[10] Z. Zhu, K. Lin, A. K. Jain, and J. Zhou, "Transfer learning in deep reinforcement learning: A survey," *IEEE Transactions on Pattern Analysis and Machine Intelligence,* 2023.

[11] H. Johra, M. Schaffer, G. Chaudhary, H. S. Kazmi, J. Le Dréau, and S. Petersen, "What metrics does the building energy performance community use to compare dynamic models?," in *IBPSA Building Simulation 2023*, 2023: International Building Performance Simulation Association, pp. 861-868.

[12] E. Lee and W. Rhee, "Individualized short-term electric load forecasting with deep neural network based transfer learning and meta learning," *IEEE Access,* vol. 9, pp. 15413-15425, 2021.

[13] Y. Ding, C. Huang, K. Liu, P. Li, and W. You, "Short-term forecasting of building cooling load based on data integrity judgment and feature transfer," *Energy and Buildings,* vol. 283, p. 112826, 2023.

[14] Y. Lu, Z. Tian, R. Zhou, and W. Liu, "A general transfer learning-based framework for thermal load prediction in regional energy system," *Energy,* vol. 217, p. 119322, 2021.

[15] C. Peng, Y. Tao, Z. Chen, Y. Zhang, and X. Sun, "Multi-source transfer learning guided ensemble LSTM for building multi-load forecasting," *Expert Systems with Applications,* vol. 202, p. 117194, 2022.

[16] K. Li, B. Wei, Q. Tang, and Y. Liu, "A data-efficient building electricity load forecasting method based on maximum mean discrepancy and improved TrAdaBoost algorithm," *Energies,* vol. 15, no. 23, p. 8780, 2022.

[17] Y. Jin, M. A. Acquah, M. Seo, and S. Han, "Short-term electric load prediction using transfer learning with interval estimate adjustment," *Energy and Buildings,* vol. 258, p. 111846, 2022.

[18] G. Mosaico, M. Saviozzi, F. Silvestro, A. Bagnasco, and A. Vinci, "Simplified state space building energy model and transfer learning based occupancy estimation for HVAC optimal control," in *2019 IEEE 5th international forum on research and technology for society and industry (RTSI)*, 2019: IEEE, pp. 353-358.

[19] P. Leeraksakiat and W. Pora, "Occupancy forecasting using lstm neural network and transfer learning," in *2020 17th International Conference on Electrical Engineering/Electronics,*

Computer, Telecommunications and Information Technology (ECTI-CON), 2020: IEEE, pp. 470-473.
[20] M. Weber, C. Doblander, and P. Mandl, "Towards the detection of building occupancy with synthetic environmental data," *arXiv preprint arXiv:2010.04209,* 2020.
[21] A. Anjomshoaa and E. Curry, "Transfer learning in smart environments," *Machine Learning and Knowledge Extraction,* vol. 3, no. 2, pp. 318-332, 2021.
[22] B. Pardamean, H. H. Muljo, T. W. Cenggoro, B. J. Chandra, and R. Rahutomo, "Using transfer learning for smart building management system," *Journal of Big Data,* vol. 6, pp. 1-12, 2019.
[23] M. Khalil, S. McGough, Z. Pourmirza, M. Pazhoohesh, and S. Walker, "Transfer learning approach for occupancy prediction in smart buildings," in *2021 12th International renewable engineering conference (IREC)*, 2021: IEEE, pp. 1-6.
[24] W.-H. Chen, P.-C. Cho, and Y.-L. Jiang, "Activity Recognition Using Transfer Learning," *Sensors & Materials,* vol. 29, 2017.
[25] K. H. Andersen, M. Schaffer, H. Johra, A. Marszal-Pomianowska, W. O'Brien, and P. K. Heiselberg, "Evaluation of Various Algorithms' Performance in Supervised Binary Classification for Occupant Detection Using a Dataset from a Residential Building," 2023.
[26] G. Pinto, Z. Wang, A. Roy, T. Hong, and A. Capozzoli, "Transfer learning for smart buildings: A critical review of algorithms, applications, and future perspectives," *Advances in Applied Energy,* vol. 5, p. 100084, 2022.
[27] P. Lissa, M. Schukat, M. Keane, and E. Barrett, "Transfer learning applied to DRL-Based heat pump control to leverage microgrid energy efficiency," *Smart Energy,* vol. 3, p. 100044, 2021.
[28] D. Coraci, S. Brandi, T. Hong, and A. Capozzoli, "An innovative heterogeneous transfer learning framework to enhance the scalability of deep reinforcement learning controllers in buildings with integrated energy systems," in *Building Simulation*, 2024, vol. 17, no. 5: Springer, pp. 739-770.
[29] D. Coraci, S. Brandi, T. Hong, and A. Capozzoli, "Online transfer learning strategy for enhancing the scalability and deployment of deep reinforcement learning control in smart buildings," *Applied Energy,* vol. 333, p. 120598, 2023.
[30] Y. Chen, Y. Ye, J. Liu, L. Zhang, W. Li, and S. Mohtaram, "Machine learning approach to predict building thermal load considering Feature variable dimensions: an office building case study," *Buildings,* vol. 13, no. 2, p. 312, 2023.
[31] W. Li *et al.*, "A perspective survey on deep transfer learning for fault diagnosis in industrial scenarios: Theories, applications and challenges," *Mechanical Systems and Signal Processing,* vol. 167, p. 108487, 2022.
[32] B. Yang, Y. Lei, S. Xu, and C.-G. Lee, "An optimal transport-embedded similarity measure for diagnostic knowledge transferability analytics across machines," *IEEE Transactions on Industrial Electronics,* vol. 69, no. 7, pp. 7372-7382, 2021.
[33] G. Li, L. Chen, J. Liu, and X. Fang, "Comparative study on deep transfer learning strategies for cross-system and cross-operation-condition building energy systems fault diagnosis," *Energy,* vol. 263, p. 125943, 2023.
[34] H. Li, T. Hong, S. H. Lee, and M. Sofos, "System-level key performance indicators for building performance evaluation," *Energy and Buildings,* vol. 209, p. 109703, 2020.
[35] H. H. Hosamo, P. R. Svennevig, K. Svidt, D. Han, and H. K. Nielsen, "A Digital Twin predictive maintenance framework of air handling units based on automatic fault detection and diagnostics," *Energy and Buildings,* vol. 261, p. 111988, 2022.
[36] *EN 13306: 2017 Maintenance - Maintenance terminology*, E. C. f. Standardization, 2017.
[37] *EN 15341:2019 Maintenance - Maintenance Key Performance Indicators*, E. C. f. Standardization, 2019.
[38] Z. Chen *et al.*, "A review of data-driven fault detection and diagnostics for building HVAC systems," *Applied Energy,* vol. 339, p. 121030, 2023.
[39] H. Li *et al.*, "Data-driven key performance indicators and datasets for building energy flexibility: A review and perspectives," *Applied Energy,* vol. 343, p. 121217, 2023.

[40] Z. L. Chen *et al.*, "A review of data-driven fault detection and diagnostics for building HVAC systems," (in English), *Appl. Energy,* Review vol. 339, p. 18, Jun 2023, Art no. 121030, doi: 10.1016/j.apenergy.2023.121030.

[41] J. Braun. "RP-1043 -- Fault Detection And Diagnostic (FDD) Requirements And Evaluation Tools For Chillers." ASHRAE. https://www.techstreet.com/standards/rp-1043-fault-detection-and-diagnostic-fdd-requirements-and-evaluation-tools-for-chillers?product_id=1716217 (accessed.

[42] J. Wen. "RP-1312 -- Tools for Evaluating Fault Detection and Diagnostic Methods for Air-Handling Units." ASHRAE. https://www.techstreet.com/standards/rp-1312-tools-for-evaluating-fault-detection-and-diagnostic-methods-for-air-handling-units?product_id=1833299. (accessed.

[43] J. Granderson, G. J. Lin, A. Harding, P. Im, and Y. Chen, "Building fault detection data to aid diagnostic algorithm creation and performance testing," (in English), *Sci. Data,* Article; Data Paper vol. 7, no. 1, p. 14, Feb 2020, Art no. 65, doi: 10.1038/s41597-020-0398-6.

[44] J. Granderson *et al.*, "A labeled dataset for building HVAC systems operating in faulted and fault-free states," (in English), *Sci. Data,* Article; Data Paper vol. 10, no. 1, p. 13, Jun 2023, Art no. 342, doi: 10.1038/s41597-023-02197-w.

[45] S. Frank, G. J. Lin, X. Jjn, R. Singla, A. Farthing, and J. Granderson, "A performance evaluation framework for building fault detection and diagnosis algorithms," (in English), *Energy Build.,* Article vol. 192, pp. 84-92, Jun 2019, doi: 10.1016/j.enbuild.2019.03.024.

[46] G. J. Lin, H. Kramer, V. Nibler, E. Crowe, and J. Granderson, "Building Analytics Tool Deployment at Scale: Benefits, Costs, and Deployment Practices," (in English), *Energies,* Article vol. 15, no. 13, p. 17, Jul 2022, Art no. 4858, doi: 10.3390/en15134858.

[47] J. Granderson, Lin, Guanjing, Singla, Rupam, Mayhorn, Ebony, Ehrlich, Paul, and Draguna Vrabie., "Commercial Fault Detection and Diagnostics Tools: What They Offer, How They Differ, and What's Still Needed," *Lawrence Berkeley National Laboratory,* 2018, doi: 10.20357/B7V88H.

[48] K. H. Andersen, S. P. Melgaard, H. Johra, A. Marszal-Pomianowska, R. L. Jensen, and P. K. Heiselberg, "Barriers and drivers for implementation of automatic fault detection and diagnosis in buildings and HVAC systems: An outlook from industry experts," (in English), *Energy Build.,* Article vol. 303, p. 20, Jan 2024, Art no. 113801, doi: 10.1016/j.enbuild.2023.113801.

[49] S. P. Melgaard, K. H. Andersen, A. Marszal-Pomianowska, R. L. Jensen, and P. K. Heiselberg, "Fault Detection and Diagnosis Encyclopedia for Building Systems: A Systematic Review," (in English), *Energies,* Review vol. 15, no. 12, p. 50, Jun 2022, Art no. 4366, doi: 10.3390/en15124366.

[50] H. Johra, "What is building energy flexibility–demand response?," 2023.

[51] H. Johra, H. Li, F. de Andrade Pereira, K. Nweye, L. Chamari, and Z. Nagy, "IEA EBC Annex 81–Data-Driven Smart Buildings: Building-to-Grid Applications," in *Proceedings of Building Simulation 2023: 18th Conference of International Building Performance Simulation Association. Shanghai, China, 4-6 September 2023*, 2023: IBPSA.

[52] F. de Andrade Pereira *et al.*, "Enabling portable demand flexibility control applications in virtual and real buildings," *Journal of Building Engineering,* vol. 86, p. 108645, 2024.

[53] F. de Andrade Pereira *et al.*, "A Semantics-Driven Framework to Enable Demand Flexibility Control Applications in Real Buildings," *Available at SSRN 4819071*.

[54] G. Chaudhary, H. Johra, L. Georges, and B. Austbø, "pymodconn: A python package for developing modular sequence-to-sequence control-oriented deep neural networks," *SoftwareX,* vol. 24, p. 101599, 2023.

[55] G. Chaudhary, H. Johra, L. Georges, and B. Austbø, "Predicting the performance of hybrid ventilation in buildings using a multivariate attention-based biLSTM Encoder–Decoder," in *Journal of Physics: Conference Series*, 2023, vol. 2654, no. 1: IOP Publishing, p. 012057.

[56] G. Chaudhary, H. Johra, L. Georges, and B. Austbø, "Synconn_build: A python based synthetic dataset generator for testing and validating control-oriented neural networks for building dynamics prediction," *MethodsX,* vol. 11, p. 102464, 2023.

[57] H. Li and T. Hong, "A semantic ontology for representing and quantifying energy flexibility of buildings," *Advances in Applied Energy,* vol. 8, p. 100113, 2022.

ANNEX 81

EBC

Energy in Building and Communities Programme

www.iea-ebc.org